A Study of the Radial and Azimuthal Gas Distribution in Massive Galaxy Clusters

A Study of the Radial and Azimuthal Gas Distribution in Massive Galaxy Clusters

A dissertation presented

by

Daniyar Rashidovich Nurgaliev

to

The Department of Physics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Physics

Harvard University

Cambridge, Massachusetts

November 2013

Thesis advisor

Christopher W. Stubbs

Author

Daniyar Rashidovich Nurgaliev

A Study of the Radial and Azimuthal Gas Distribution in Massive Galaxy Clusters

Abstract

Clusters of galaxies are particularly interesting astrophysical systems, are the largest bound structures in the Universe, and contain fair sample of cosmic ingredients. Studies of cluster abundance as a function of mass and redshift were critical in establishing the standard model of cosmology. This dissertation presents results from X-ray imaging of massive distant ($M > 10^{14} M_{\odot}$, $0.3 < z < 1.2$) clusters, found via X-ray emission or Sunyaev-Zeldovich effect. This is the world's largest sample of massive galaxy clusters. We explore the radial and azimuthal profiles of the X-ray emitting gas and show that clusters are self-similar objects: their internal structure is largely independent of the cluster's mass or redshift, and the fractions of different types of clusters does not change with redshift. We also present a new statistical technique for measuring a cluster's deviations from a perfect axisymmetric shape, which is especially useful in the case of low photon count observations of distant clusters.

Contents

Relation to Previously Published Work

Chapter 2 is set to appear in the following paper:

"Universality of the radial profiles of X-ray emitting gas in massive galaxy clusters", D. Nurgaliev, M. McDonald, B. A. Benson, C. W. Stubbs, V. Vikhlinin, The SPT Collaboration

Chapter 3 was released in the following publication:

"Robust Quantification of Galaxy Cluster Morphology Using Asymmetry and Central Concentration", D. Nurgaliev, M. McDonald, B. A. Benson, E. D. Miller, C. W. Stubbs, V. Vikhlinin, *arXiv:1309.7044, Ap. J.* **779**, 2013

Chapter 4 is set to appear in the following paper:

"Testing for morphological bias between X-ray and SZ-detected galaxy clusters", D. Nurgaliev, M. McDonald, B. A. Benson, C. W. Stubbs, V. Vikhlinin, The SPT Collaboration

Acknowledgments

I would like to thank Christopher Stubbs for giving me the opportunity to work on this research project, as well as for his constant support and advice over the years.

I would like to give special acknowledgements to Alexey Vikhlinin (who is on my thesis committee) and Michael McDonald who mentored me during my time in Chris Stubbs's lab, and whose support and ideas were important for completion of this work.

I also wish to express gratitude to the third member of my thesis committee, Alyssa Goodman.

The completion of this work would have been impossible without the teaching, support and friendship of the colleagues with whom I worked directly, most of all Gautham Narayan, Jonathan Ruel, Brian Stalder and Matt Bayliss. I also credit the help and advice of many other members of the SPT team, most notably Bradford Benson, Eric Miller, and Tony Stark.

Finally, I would like to thank Alice Xiang who helped me overcome some of my language difficulties and restored hundreds of missed "the" articles in my writing.

I acknowledge support by the National Science Foundation grant AST-1009012.

Dedicated to the proletarians of the world

Chapter 1

Introduction

Galaxy clusters were first identified as the agglomeration of galaxies more than 200 years ago (Biviano, 2000). Over time it became evident that galaxy clusters are much more than just an aggregation of galaxies. In fact, only a few percent of cluster mass is in the form of stars and galaxies (Gonzalez et al., 2007), the rest being hot inter cluster gas (10-15%) and dark matter ($\sim$ 85 %). At the scale of galaxy clusters ($\sim$ Mpc), gravity is the dominant force defining the distribution of matter. The dark matter component, therefore, is the major component that drives cluster evolution. Both intergalactic gas and galaxies follow the gravitational potential created by dark matter particles, but have little, if any, influence on the dark matter itself.

Even though the observable part of galaxy clusters, namely the inter cluster medium (ICM) and galaxies constitute only a small part of cluster mass, they are perfect tracers of dark matter distributions, and therefore are rich sources of information about dark matter and the behavior of a self-gravitating system.

In the rest of the introduction we will first discuss how the dark matter component evolves

from an initially homogeneous universe into the present day large scale structure, where galaxy clusters are the largest gravitationally bound systems that demonstrate extreme density contrast compared to the mean density of the universe. Then we will discuss the observation of galaxy clusters, the methods for estimating their mass and, finally, the details of their internal (both radial and azimuthal) structure.

1.1 Formation of Galaxy clusters

1.1.1 Growth of large scale structure, gravitational collapse

A system of self-gravitating particles is unstable – infinitesimally small fluctuations in their density grow over time producing larger and larger inhomogeneities and clumps of matter. In the early stages – when the density fluctuations are small – their evolution can be described by the following linearized equation, which takes the simplest form in the case of non-relativistic collisionless particles which is the currently accepted model of dark matter

$$\frac{\partial^2 \delta}{\partial t^2} + \frac{2\dot{a}}{a}\frac{\partial \delta}{\partial t} = 4\pi G \overline{\rho} \delta. \tag{1.1}$$

Since this equation doesn't contain any derivative with respect to the x-coordinate, the overdensities at different x simply grow independently. The dynamics of the growth varies slightly depending on the cosmological parameters, but in all cases the growth factor is close to the scale factor of the universe, i.e.

$$\delta(x,t) = a(t)\delta_0. \tag{1.2}$$

The linear approximation made in eq. 1.2 breaks down at the values of $\delta(x,t)$ of order

unity and the linear growth changes to the gravitational collapse and decoupling of the given density perturbation from the growth of structure in the Universe. The highest (and rarest) peaks of primordial density fluctuations collapse first, giving rise to galaxy clusters that represent the biggest gravitationally bound objects in the Universe. The collapse gradually spreads to the neighboring regions with lower overdensity, which corresponds to the gradual growth of a dark matter halo by further accretion of matter.

The particles within a dark matter halo equilibrate through their gravitational interactions in the process called "violent relaxation". The resulting mean overdensity of the virialized sphere is approximately 200 times the critical density of the universe at the moment of virialization.

Due to the power-law spectrum of primordial density fluctuations, the growth-of-structure process is initially scale-free and therefore produces a population of objects – galaxy clusters – which are essentially scaled versions of one another. Therefore, they have similar shapes and all their global properties can be related to their mass through power-law dependencies called "scaling relations".

The theoretical slopes of scaling relations are the most important consequences of self-similar theories. At the same time, scaling relations are particularly relevant to relating galaxy clusters to cosmological parameters. As a result, scaling relations are studied extensively from theoretical, numerical and observational points of view. Observations confirm strong correlations of global cluster properties such as total mass, gas mass, luminosity, temperature etc. but also show some deviations fro the theoretically motivated exponents.

1.2 Observations of Galaxy clusters

The formation of galaxy clusters is defined by their dark matter component as it comprises the dominant $\sim 85\%$ of total cluster mass. Thus, cluster formation can be approximated as the dissipationsless gravitational collapse of dark matter. Following the dark matter's gravitational potential, the baryons are heated up and form stars and galaxies. Due to the inefficiency of star formation, only $\sim 10\%$ of the baryonic matter is converted into stars and galaxies. The remaining $\sim 90\%$ fills up the intergalactic space forming the Inter Cluster Medium (ICM) – hot plasma emitting in X-rays. As the dark matter is not directly observable, the two baryonic components, galaxies and the ICM, provide the majority of information about galaxy clusters.

The three main methods of galaxy cluster detection are based on 1) optical detection of member galaxies by their overdensity and color properties, 2) X-ray detection due to the thermal bremsstrahlung from the hot ICM and 3) detection based on the Sunyaev-Zeldovich effect (Sunyaev & Zeldovich, 1972)– the change of the spectrum of CMB radiation scattered off the ICM.

1.2.1 Optical

Historically galaxy clusters were identified as regions in the sky with an overdensity of galaxies. First systematic galaxy cluster catalog was compiled by Abell (1958), who established a number of criteria on the number and magnitudes of the galaxies within a certain solid angle which corresponds to a physical radius of approximately 1.5Mpc. Abell's criteria turned out to be very effective and laid out a foundation for modern optical detection techniques.

Nowadays galaxy clusters are detected in optical surveys by searching for over-densities of galaxies with known color properties. Compared to other wavelengths, optical cluster surveys provide the advantage of a high S/N ratio and a wide field of view. Because of a greater depth, these surveys give the largest yield in terms of cluster detections per square degree. Galaxy clusters detected in large optical surveys like SDSS make the biggest galaxy cluster samples (Koester et al., 2007), so optical detection remains an important method of discovery of galaxy clusters .

Optical surveys are also essential to provide photometric (and spectroscopic) redshift information which is harder (X-ray) or impossible (SZ) to obtain from other wavelengths. Optical confirmation is a necessary step for all SZ-detected clusters.

The masses of galaxy clusters correlate with their "richness" – the number of galaxies above some luminosity threshold. The other way to estimate cluster's mass by optical observation is measuring the line-of-sight velocity dispersion of member galaxies. Although the velocity dispersion – mass relation has a high intrinsic uncertainty, it directly probes the gravitational potential providing a mass estimate which, unlike X-ray and SZ-masses, is uncorrelated with the properties of the ICM.

An even more direct way of measuring cluster masses is by means of weak lensing. Currently, this method is not as sensitive as others, but can eventually serve as a direct calibration tool for other methods (High et al., 2012; Hoekstra, 2007).

1.2.2 X-rays

Galaxy clusters demonstrate an extended X-ray emission which is used in surveys such as the Rosat All-Sky Survey (Bohringer et al., 2000; Cruddace et al., 2002) and serendipitous

surveys such as 400 square degrees survey (Burenin et al., 2007). Also, X-ray observations can be used to follow up SZ-detected clusters, since they provide more information than optical observations (e.g. Planck's X-ray follow-up program Planck Collaboration et al. (2011, 2012, 2013c)).

X-ray observations are currently the most precise source of information from the galaxy clusters. Clusters are visible in X-rays due to bremsstrahlung radiation from the ICM plasma. The ICM makes $\sim 90\%$ of cluster baryonic mass and is much more regular than the galaxy distribution. The ICM in dynamically relaxed systems appears to be in approximate hydrostatic equilibrium, which allows for direct measurement of total mass profiles. Hydrostatic masses from the local relaxed sample are used to fit scaling relations which are then transferred to arbitrary z by the assumptions of self-similar evolution and comparisons with simulations.

Another unique feature of X-ray observations of galaxy clusters is that the element abundances in the ICM can be determined from X-ray spectra.

1.2.3 Sunyaev-Zeldovich effect

The hot gas in galaxy clusters scatters the cold CMB photons, on average increasing their energy. As the result, the spectrum of the CMB in the direction of the cluster changes. The net number of CMB photons below 220GHz is decreased, and the missing photons enhance the CMB signal above 220GHz. Although the SZ-effect was predicted ~ 40 years ago (Sunyaev & Zeldovich, 1972), only recently has it become a mainstream method of cluster detection with 3 big surveys (SPT, ACT, Planck) recently finished. However, the advancement of microwave detectors needed to observe the SZ effect is so fast that the SZ

effect may soon become the primary method of cluster detection.

The most attractive property of SZ observations for cosmological studies is the independence of the SZ signal on redshift. Thus, an SZ survey finds clusters out to arbitrarily high redshift therefore probing the large scale structure in the early Universe and allowing the studies of the clusters in the epoch of their formation around $z = 2$. On the other hand, the independence of redshift property makes follow up observations in X-ray or optical wavelengths necessary.

Another attractive feature of the SZ signal is its proportionality to the integral of the electron concentration times the temperature over the volume of the cluster. In other words, the SZ signal is proportional to the total thermal energy of the cluster, potentially making it the lowest-scatter mass proxy.

One of the disadvantages of SZ detection is its contamination with radio sources. This problem, however, may be resolved by multiwavelength microwave observations and fitting the spectrum to the expected SZE spectrum.

1.3 Determination of cluster mass

1.3.1 Why mass is important

The masses of clusters is mostly interesting from the cosmological point of view since the cluster abundance as a function of mass and redshift can provide some of the most powerful constraints on cosmological models. From the astrophysical point of view, all of the global properties of the baryonic component of clusters such as temperature or luminosity are linked to the depth of the gravitational well and therefore to mass. Also, the mass of a cluster sets

the energy scale to be compared with energy scales of other astrophysical processes such as radiative cooling rates and stars and AGN feedback.

Since clusters don't have a well defined boundary, the definition of their spatial extent and total mass is ambiguous. The common way to define cluster's radius and mass is through a sphere enclosing a certain mean overdensity Δ compared to the critical density of the Universe, or in other words by solving implicit equation.

$$M_{\Delta}(< R_{\Delta}) = \Delta \frac{4}{3} \pi \rho_c R_{\Delta}^3 \tag{1.3}$$

The usual choices for Δ are 200 which makes R_{Δ} close to virial radius, and 500, which is more convenient for X-ray observations.

1.3.2 Hydrostatic mass

For local galaxy clusters both local density and temperature can be measured with good precision. To find the radial temperature distribution, a cluster is split into a series of "onion skins", with a different temperature in each of them. Having this 3-D temperature map, one can compute the expected X-ray spectrum at different projected radii and fit to the observed spectrum. The same can be done for density. Having the temperature and density dependence on radius, the enclosed mass is given by the equation of hydrostatic equilibrium

$$M_{HE}(< r) \; = \; -\frac{r k_B T(r)}{G \mu m_p} \left[\frac{d \ln \rho_{\rm g}(r)}{d \ln r} + \frac{d \ln T(r)}{d \ln r} \right] . \tag{1.4}$$

where μ is the mean molecular weight, k_B is Boltzmann constant and m_p is the proton mass.

The hydrostatic equilibrium equation provides the most direct and precise method of measuring masses for relaxed local clusters and serves as a calibration tool for all other

mass estimates. Therefore, deviations from hydrostatic equilibrium are likely to account for the largest systematic errors in X-ray measurements of cluster masses.

Both simulations (Nagai et al., 2007) and observational comparison with weak lensing masses (Mahdavi et al., 2008) show that hydrostatic equation underestimates total mass by $\sim$ 10-20% at R_{500}. This is most probably the result of non-thermal pressure support from turbulent gas motion and magnetic fields (Nagai et al., 2007). Further comparison of hydrostatic and weak masses is a subject of active research (e.g. Mahdavi et al., 2013; Zhang et al., 2010; Israel et al., 2010)

1.3.3 Scaling relations

The regularity of the cluster population is manifested in tight correlations between their integrated observable properties such as mass, temperature, luminosity, gas content, SZ integrated signal etc., which have power-law dependences on the cluster's mass. This fact is manifested in the so-called scaling relations which may be expressed in the following form:

$$\log O = \log A + B \log M/M_0 + C \log E(z)/E(z_0) + \mathcal{N}(0, D^2), \tag{1.5}$$

where A defines the normalization of the observable, B is usually referred to as a slope of scaling relation, C characterizes redshift evolution of the relation, and D characterizes the scatter of an observable O around the relation.

The hierarchical model of cluster formation (Kaiser, 1986) provides the theoretical predictions of the exponents B of various cluster observables. However, the observed exponents of these relations deviate from those predicted theoretically. Some of the observed exponents of scaling relations, such as $M - T_X$ and $M - Y_X$ are almost consistent with the theoretical ones (Vikhlinin et al., 2009; Benson et al., 2014) whereas the others, such as the $M - L_X$

relation, show significant departure from self-similarity. The $M - L_X$ relation which has an observed slope of ~ 3 instead of theoretically reasoned 2 (Maughan et al., 2012) is the most notable departure from self-similarity.

The $M - T_X$ scaling relation is particularly useful and well-studied, because it can be applied even when the cluster is not spatially resolved. Additionally, it is widely used because it is easy to use - one only needs to measure the X-ray spectroscopic temperature. The $M - Y_X$ relation has the smallest scatter, and therefore provides the most precise mass estimates. The $M - Y_{SZ}$ relation is also expected to have a very small scatter. However, measuring Y_{SZ} with present-day SZ-telescopes is a challenging and instrument-dependent task (Saliwanchik et al., 2013) due to insufficient resolution of present day SZ telescopes. As a consequence, Y_X is not well calibrated yet and some of SZ-telescope collaborations prefer using instrument- and depth-dependent signal-to-noise ratio as a mass proxy (Vanderlinde et al., 2010).

1.3.4 Using multiple mass proxies, and corrections for dynamical state

It is tempting to combine all mass proxies into a single mass estimator which would have lower scatter. Constructing such an estimator, though, requires a good understanding of the covariances between different mass observables. Some of the work in that direction has been done both for simulated (e.g. Noh & Cohn, 2012; Angulo et al., 2012) and observed clusters (e.g. Rozo et al., 2009), but the results are not yet widely accepted. One of the reasons why multiple mass observables are hard to combine into a single mass proxy is that the quality of measurements of various observables may vary substantially from one cluster to another

and from one dataset to another dataset. Apparently, these variations would have to be taken into account while combining different measurements into a single mass proxy.

Perhaps the only method of combining different mass observables which is extremely successful is $Y_X = M_{gas}T_X$ (Kravtsov et al., 2006). It elegantly avoids the 2 complications for combining various mass proxies outlined in the previous paragraph. First, instead of studying covariance between M_{gas} and T_X, the use of their product as a mass proxy is motivated theoretically as it approximates the total thermal energy of the cluster. Second, both M_{gas} and T_X are computed from the same X-ray observation, and therefore cannot significantly differ in quality. Y_X is the mass proxy with the lowest known scatter. One of the explanations of the low intrinsic scatter in the $M - Y_X$ relation is the anticorrelation of deviations ΔM_{gas} and ΔT_X between true and measured observables M_{gas} and T_X (Kravtsov et al., 2006).

Another way to decrease the scatter around scaling relations is to explicitly take into account the dynamical state of the cluster. Simulations show that recent mergers deviate from scaling relations (Krause et al., 2012; Ventimiglia et al., 2008; Poole et al., 2007; O'Hara et al., 2006). For example, Vikhlinin et al. (2009) includes a 17% correction to mergers in the $M - T_X$ scaling relation, which was supported by Kravtsov et al. (2006) simulations.

Finally there are multiple studies that combine multiwavelength observations to build complex models of individual systems. The popular choices of observations to combine are X-ray and SZ because X-rays can constrain the inner parts of the cluster and SZ measures the outer parts (Bonamente et al., 2012; Ameglio et al., 2009, 2007; Bonamente et al., 2004); X-ray and lensing observations (Mahdavi et al., 2013; Konrad et al., 2013) or all of them (Morandi et al., 2012; Sereno, 2007; Mahdavi et al., 2007; Reblinsky & Bartelmann, 1999).

1.4 Internal structure of galaxy clusters

1.4.1 Radial profiles of density, temperature, pressure and entropy

In the simplest, spherically symmetric approximation for galaxy clusters (which also proves to be valid with good approximation (Buote & Humphrey, 2012b), at least on average) all the information about their internal structure is contained in the radial profiles of their total mass (which describes the cluster's dark matter component); temperature, density, pressure and entropy (which describe the ICM component); and radial profiles of galaxy density and velocity dispersion (which describe the galaxy component). Studies of these profiles provide insights into cluster formation histories, thermodynamics and astrophysical properties in ICM, and are important for associating a cluster's integrated observables such as Y_X and Y_{SZ} with cluster mass.

For example, pressure profiles are directly related to total cluster mass by hydrostatic equilibrium and to integrated SZ-signal which is effectively the integral of pressure over the volume of the cluster. The vast, low pressure cluster and group outskirts introduce strong corrections to the secondary CMB anisotropy at hi-ℓ (Shaw et al., 2010). Entropy profiles are especially important for characterizing the heating and cooling processes in ICM and the accretion phenomena. (McNamara & Nulsen, 2007; Voit, 2005)

1.4.2 Modelling of radial profiles

Dark matter

The distribution of matter is one of the most fundamental questions in the study of bound systems such as galaxies or galaxy clusters. In the spherically symmetric approximation (which we usually have to make due to the lack of observational data for more complex models) of galaxy clusters, the radial density profile carries all of the cluster's matter distribution information. The precise functional form of the radial profile is of great interest for understanding the nature of dark matter, the interaction between dark matter and baryonic matter, the processes of accretion from outside the virial radius, and the equilibration processes after mergers. Since early studies on the self-similar collapse of primordial density fluctuations (Fillmore & Goldreich, 1984; Bertchinger, 1985), it has been recognized that virialized halos are described by power-law density profiles. At first, the models were described by a single power law. As N-body techniques improved, two different power laws have been observed respectively in the inner and the outer parts of the dark matter halos. Navarro et al. (1995) published their famous mass density profile (NFW)

$$\rho(r) = \rho_0 \frac{1}{r} \frac{1}{(1 + r/r_s)^2} \tag{1.6}$$

which has become the de facto standard in empirical descriptions of dark matter halos. Despite the popularity of the NFW profile, there were arguments in the literature (e.g. Moore et al., 1999) that an inner profile of $\rho \sim r^{-1.5}$ may be more appropriate. Therefore, many authors have resorted to the generalized NFW profile which leaves the inner power law exponent as a free parameter and gives additional freedom to the second factor in the

equation

$$\rho(r) = \rho_0 \frac{1}{r^\alpha} \frac{1}{(1 + (r/r_s)^\gamma)^{\frac{3-\alpha}{\gamma}}}. \tag{1.7}$$

Another approach to the description of matter halos was recently proposed by Navarro et al. (2004) and defined by the equation:

$$\rho(r) \sim \exp(-Ar^\alpha) \tag{1.8}$$

which, unlike NFW-like profiles has a constant change of slope in log-log plot. It is interesting that Navarro's (2004) model happens to be a revival of the model describing the density of a spherical stellar system introduced back in 1963! (Einasto, 1965). Another interesting feature of the Einasto model is that it does not have a characteristic radius at which the power law changes. Instead, the logarithmic slope varies continuously with radius. There are indications that the Einasto model fits simulated dark matter halos better than the widely used NFW model.

Obviously, it is hard to measure the dark matter profile experimentally. However, it can be done for some nearby clusters using the assumption of hydrostatic equilibrium, e.g. Pointecouteau et al. (2004); Pratt & Arnaud (2002) have done it for A478, A2029 and A1413 respectively and found that GNFW, rather than NFW provides a better fit.

X-ray emitting gas

The distribution of baryons in galaxy clusters differs from the distribution of dark matter. Most notably, the majority of clusters have a flat density baryonic core. The gas density profiles of such clusters are reasonably well described by the so-called beta-model (or King model)

$$\rho = \rho_0 \frac{1}{(1 + (r/r_c)^2)^{3\beta/2}}. \tag{1.9}$$

The beta model has a simple functional form and relies on theoretical models of isothermal dark matter particles, which can explain its popularity. However, it fails to account for two features observed in real clusters: cuspy X-ray brightness profiles associated with cool core clusters, and further steepening of the logarithmic density slope in the cluster outskirts. These shortcomings served as a motivation for the introduction of the generalized beta model (Vikhlinin2009):

$$\rho = \rho_0 \, \frac{(r/r_c)^{-\alpha/2}}{(1+r^2/r_c^2)^{3\beta/2-\alpha/4}} \frac{1}{(1+r^\gamma/{r_s}^\gamma)^{\varepsilon/2\gamma}}, \qquad \gamma = 3. \tag{1.10}$$

(The original Vikhlinin model contained a second term that provides additional freedom in modeling cool cores, but as the model as presented in eq. 1.10 fits even cool core clusters well, it has been widely used without the second term).

Overall, all these functional forms look a bit mysterious and it is not easy to develop an intuition about them. For example, looking at the eq. 1.7, it is hard to say what exactly is controlled by the parameter γ. All these functional forms will, however, become much more transparent in our unified radial profiles framework.

A unified framework

As we can see above, there are many analytic expressions that are used to fit radial density (and pressure) profiles for both the dark matter and baryonic component, starting from the famous NFW profile and ending with the Bulbul et al. (2010) profile. They use a variety of analytic functions, which at first sight seem a little arbitrary. However, we will show that these choices look more reasonable when we think of radial profiles in terms of their logarithmic derivative alpha

$$\alpha = -\frac{d\ln\rho}{d\ln r}. \tag{1.11}$$

Instead of discussing radial profiles in their original coordinates ρ, r, we will discuss their meaning in the coordinates (α, ξ), where $\xi = \ln r/r_0$, and r_0 is a characteristic size of the object. This approach, which is fully developed in Sec. 2.4.1 proves to be very productive in reconciling many functional forms of radial profiles discussed in literature.

1.4.3 Departures from assumed spherical symmetry

Currently all the techniques for mass estimation of galaxy clusters rely on the assumption of spherical symmetry. The advantage of this approach is a comparative simplicity of data analysis. It is believed that non-spherical effects do not dominate the error budget. However, clusters grow from initially triaxial density peaks. Both simulations (Lau et al., 2011; Kasun & Evrard, 2005; Jing & Suto, 2002) and observations (Lau et al., 2012; Kawahara, 2010) show that clusters are elliptical with an average ratio of minor to major axes of 0.7. One of the factors that makes the observation of ellipticities so difficult is that no method can reliably estimate the elongation of cluster along a given line of sight. Therefore, we can only resort to quantifying the ellipticities of clusters projections onto the plane of the sky.

In principle, multiwavelength observations may constrain 3-D shapes, at least for some clusters. Since the SZ signal is sensitive to density and the X-ray emission is sensitive to density squared, a combination of X-ray and SZ observations can provide constraints on cluster shapes. Both simple model-independent constraints on the line-of-sight elongation (Mahdavi & Chang, 2011) and complex models of triaxial halos of arbitrary orientation (Morandi et al., 2012; Sereno, 2007; Mahdavi et al., 2007) were built. One of the important results that can be obtained from combined X-ray and SZ observations apart from cluster shapes are constraints on the ICM gas clumping, which is necessary for understanding the

systematics in X-ray derived masses.

Velocity dispersion is another candidate for measuring cluster shapes. Since the velocity ellipsoid is aligned with the cluster ellipsoid to a high degree (Kasun & Evrard, 2005) one can associate line-of-sight velocity information of a member galaxy with its actual position (not redshift!) z along line of sight, and the positional information on the sky with the corresponding projected position in the cluster x, y.

The shape of a galaxy cluster is sensitive to the details of baryonic physics, and can be used to obtain constraints on gas physics: star formation, thermal feedback due to supernovae, AGN and UV heating, metal enrichment and metallicity-dependent cooling (Lau et al., 2011, 2012).

In defense of the currently widely applied spherical approximation of cluster shapes, I would like to note that nonsphericity has only a small influence on X-ray observables derived in the spherical approximation and the hydrostatic equilibrium equation. This is not surprising – after averaging over a spherical shell, ellipticity may only introduce a second order correction to averaged quantities. For example. Pratt & Arnaud (2002) showed that a relatively high ellipticity of A1413, ($\epsilon = 1.4$ in projection), has negligible impact on the derived temperature and mass profiles. Nevertheless, Buote & Humphrey (2012a) argues that with the development of precision cosmology, the effects of additional scatter and biases due to clusters' nonsphericity will become increasingly important.

Buote & Humphrey (2012a,b) have performed an in-depth analysis of the effects of spherical averaging of elliptical galaxy clusters. They concluded that the resulting scatter in Y_{SZ}, M_{gas}, Y_X is about 10% which is smaller, but comparable to the typical uncertainty of Y_{SZ}, M_{gas}, Y_X of 15-20%. The orientation-averaged bias, however, is below 1% and therefore

not significant compared to other sources of systematic error.

There are lots of finer substructures in galaxy clusters that can reveal a lot about a particular cluster's history. For example, sharp central peaks in the X-ray emission are associated with cool cores and the long periods of isolation needed for cool cores to develop. Spiral structures suggest that the cluster has recently undergone a minor off-axis merger. Complex morphologies that include shocks, clumps, filaments and multiple peaks of X-ray surface brightness are indicative of the ongoing merging activity between two or more systems. Since "merginess" is an important characteristic of clusters which is likely to have the strongest influence on scaling relations (Sec. 1.3.3), a considerable effort has been undertaken to quantify "merginess" based on X-ray surface brightness. Some of these efforts are described in Chapter 3.

Chapter 2

Universality of the radial profiles of X-ray emitting gas in massive galaxy clusters

2.1 Introduction

Galaxy clusters are the largest gravitationally bound objects in the universe, and therefore are ideal laboratories for studies of gravitational physics at the largest possible scale. The main source of information about the internal structure of galaxy clusters is the X-ray emission from hot plasma (the Inter Cluster Medium, or ICM) that fills up the intergalactic space in clusters. The shapes of both dark matter and baryonic components of galaxy clusters are influenced by many processes happening over cosmic time: the spectrum of primordial density fluctuations, the details of density perturbation growth and gravitational collapse, the processes of hierarchical merging and virialization, the baryonic feedback in the

form of excess energy infusion by supernovae and AGNs, the radiative cooling, the hydrodynamic processes in the ICM etc. Studies of ICM properties in general, and of their density profiles in particular, can give insight into the interplay of these processes and test the assumptions of various cluster formation theories. One of the most important assumptions of such theories is self-similarity of the cluster population.

2.1.1 Self-similarity

Galaxy clusters emerge from the gravitational collapse of rare high peaks of primordial density fluctuations (Press & Schechter, 1974) which are initially scale-free (Harrison, 1970; Zeldovich, 1972). The growth factor of primordial density perturbations has complex dependence on the scale of these perturbations and on the cosmological model . However, the net result of the density fluctuation growth for length scales that are relevant for clusters is that they can still be described by a scale-less power law spectrum $P(k) \propto k^n$. A typical (rms) overdensity of linear size R is then given by the following expression

$$\langle \delta_R^2 \rangle \equiv \sigma^2(R) = \frac{1}{(2\pi)^3} \int P(k) |\tilde{W}(\mathbf{k}, R)|^2 d^3k \ \propto \ R^{-(n+3)}, \tag{2.1}$$

where $\tilde{W}(\mathbf{k}, R)$ is the Fourier transform of the top-hat window function which determines the linear scale R. The probability of fluctuations of a given overdensity and size δ_R can be computed using the introduced RMS overdensity $\sigma^2(R)$ as

$$p(\delta_R) = \frac{1}{\sqrt{2\pi}\sigma(R)} \exp\left[-\frac{\delta_R^2}{2\sigma^2(R)}\right]. \tag{2.2}$$

Since R and M are interchangeable through the relation $M = 4\pi/3\rho_{\rm m}(z)R^3$, equation (2.2) also gives the probability of finding a fluctuation of a given overdensity and mass.

As the normalization of the power spectrum grows over time (approximately proportional to the scale factor $a(t)$), so does the density contrast δ_M of any given perturbation. When it becomes comparable to unity, the fluctuation decouples from the linear growth of structure and enters the non-linear regime of gravitational collapse. When the linearly extrapolated overdensity reaches the value of $\delta\rho/\rho \simeq 1.69$, the fluctuation is thought to be completely virialized red(). The mass scale $M_{\rm NL}$ at redshift z for which a typical perturbation reaches this threshold should satisfy the equation $\sigma(M_{\rm NL}, z) \simeq 1.69$. This so-called non-linear mass scale is the only dimensional parameter in the problem of evolution of structure, and therefore all properties of clusters including their abundance and details of their internal structure can only depend on the ratio of cluster mass to $M_{\rm NL}(z)$. To see this, let's look at the equations 2.1, 2.2 again – all statistical properties of the fluctuation field smoothed at a mass scale M depend on $\sigma(M, z)$ which in the case of a power law spectrum of fluctuations can be expressed as

$$\begin{gathered} \sigma^2(M,z) \propto R^{-(n+3)} \propto M^{-(1+n/3)} \\ \sigma^2(M,z) = \sigma^2(M_{\rm NL}(z),z)\left(\frac{M}{M_{\rm NL}(z)}\right)^{-(1+n/3)} = 1.69^2\left(\frac{M}{M_{\rm NL}(z)}\right)^{-(1+n/3)} . \end{gathered} \tag{2.3}$$

The fact that $\sigma^2(M, z)$ is expressed as a simple power law in mass $\sigma^2(M, z) \propto (M/M_{\rm NL})^{-\gamma}$, with $\gamma = 1 + n/3$, has two consequences: 1) at a given redshift all overdensities of different mass scales M are composed of statistically the same combinations of fluctuations of proportionally smaller scales. 2) The redshift dependence enters eq. 2.3 only through $M_{\rm NL}$, therefore clusters formed at different redshifts are also just scaled versions of one another. In other words, the internal structure of density fluctuations is the same for cluster progenitors at all masses and redshifts, after scaling to the size of that cluster progenitor. For more elaborate theoretical models predicting the distribution of sub-clumps in clusters see e.g.

Sheth (2008); Lacey & Cole (1993).

The dependence of $M_{\rm NL}$ on the scale factor $a(t)$ is also a power law to a good approximation. This means that if we move back in time from the redshift z_0 when a particular cluster virialized to a redshift when the Universe was a factor of x smaller, we would see a distribution of matter in place of the future cluster which has exactly the same statistical properties, independent of z_0.

The virialization of the particles in a cluster is another scale-free process (Lynden-Bell, 1067) and should lead to an equilibrium state eventually independent of initial conditions with self-similar radial profiles. However, the virialization and equilibration of ICM is a long process: the substructures falling into a cluster can still be seen Gigayears after the start of a merger (Nelson et al., 2012; Poole et al., 2006; O'Hara et al., 2006). Thus the fact that the initial conditions are self-similar and independent of cluster mass and redshift, ensures that clusters remain approximately self-similar throughout the virialization process.

The overall picture that we are arriving at is that clusters are self-similar objects. Their dimensionless properties (e.g. abundance, gas content, concentration of their mass distribution and others (Kravtsov et al., 2005)) are determined by the ratio of their virial mass to the nonlinear mass scale $M_{\rm NL}$, while others (e.g. temperature, luminosity, etc (Kaiser, 1986)) have a power law dependence on their mass. This power law dependence in the form of scaling relations is a subject of intensive research (see Giodini et al., 2013, for review) due to the importance of determining cluster masses for constraining cosmological parameters. In this paper we will, however, concentrate on the self-similarity of the internal structure of galaxy clusters, in particular their radial density profiles. From the theoretical arguments presented above the radial density profiles are expected to be the same in terms of properly

scaled quantities (e.g. $r/R_{vir}, \rho/\rho_{mean}$, etc).

Some weak deviations from the self-similarity discussed above are expected even in the absence of any process that breaks self-similarity. An example of this is the dependence of the concentration of dark matter halos on mass. More details on this topic are in Kravtsov & Borgani (2012).

One of the important deviations from the spherical gravitational collapse theory is that clusters form from initially triaxial peaks of matter distribution, and due to orbital instability and preferential accretion from filaments, further develop their elliptical shape. For moderate ellipticities, however, averaging profiles in spherical shells has a negligible effect on the derived profiles (Pratt & Arnaud, 2002).

All the theoretical reasonings presented above are in agreement with N-body simulations. Radial profiles of clusters of different masses are nearly scaled versions of one another (Duffy et al., 2008), although they may have a weak dependence on $M/M_{\rm NL}$ (Gao et al., 2008). Some other cluster properties, such as f_{gas} also depend on $M/M_{\rm NL}$ only (Kravtsov et al., 2005).

For the practical application of self-similar scaling we would need to know the clusters' virial masses (which are not well defined in the case of real, not spherical collapse). Both theoretical considerations , and simulations agree that the spherical overdensity mass enclosed by R_{200} where $\rho = 200\rho_{cr}$ is a good proxy for a cluster's virial mass. However, masses determined from X-ray observations are typically M_{500} corresponding to critical overdensity of 500, because X-ray emission decreases sharply with radius, and is typically only accessible within R_{500}. We argue that using M_{500} and R_{500} for the proper scaling of clusters is almost as good as M_{vir} and R_{vir}. This is because M_{500} and R_{500} are approximately proportional

to M_{vir} and R_{vir}. The proportionality would be precise in the idealized picture of spherical collapse presented above, and remains approximately correct for the realistic growth of clusters. (One can see this is the case for the NFW model if the concentration doesn't change much for the considered range of masses).

In the case of X-ray observations of distant clusters M_{500} can only be obtained through scaling relations from the X-ray observables: T_X, M_{gas}, and $Y_X = T_X M_{gas}$ (Vikhlinin et al., 2009). M_{gas} is not suitable for testing the self-similarity of density profiles because M_{gas} is essentially determined from the same density profiles, so it would be totally unsurprising if the profiles coincided after scaling by M_{gas}. T_X is completely independent of the surface brightness, so that comparing surface brightness after scaling profiles to R_{500} computed from T_X is the best test of self-similarity. Unfortunately, $R_{500}(T_X)$ is sensitive to the dynamical state of the cluster (Kravtsov et al., 2006), and therefore has a high scatter. Finally, Y_X is known to be the best mass indicator, so we expect the profiles to be more self-similar when they are scaled by Y_X. Also, Y_X is insensitive to the cluster dynamical state, and has a much weaker dependence on surface brightness than M_{gas}.

2.1.2 Self-similarity in observed clusters

Observations confirm the similarity of cluster shapes. All clusters, after appropriate radial scaling and normalization, can be described by "the universal profiles". These universal density, pressure, and temperature profiles have been widely studied in the literature (Eckert et al., 2011; Arnaud et al., 2010; Croston et al., 2008; Arnaud et al., 2002; Neumann & Arnaud, 1999). All authors agree that clusters can only be considered self-similar outside of their core regions of approximate size $\sim 0.15 - 0.3 R_{500}$. Within the core regions

astrophysical processes, rather than gravity are most influential in shaping the gas distributions. The most striking example of deviations from self-similarity in cluster cores is the separation of the cluster population into cool-core and non-cool-core subsets. Cool core and non-cool core clusters differ in their central properties by orders of magnitude. There are also indications that the central properties (and profiles near centers) correlate with redshift (Vikhlinin et al. (2007); McDonald et al. (2013)). Evidently, both these properties break the self-similar model. On the other hand, in the cluster outskirts, where gravitational physics dominates, galaxy clusters show a remarkable regularity in their density, temperature, and pressure profiles, both in observations and simulations (Arnaud et al., 2010; Nagai et al., 2007).

Pressure profiles are self-similar and more regular than density profiles (e.g. in the cores temperature and density anticorrelate (CC vs NCC), so the scatter in pressure is smaller than in density or temperature). The deviations from self-similarity of pressure profiles are associated with the cluster's dynamical state (Arnaud et al., 2010). Pressure profiles are important for calibrating SZ scaling relations and hi-ℓ end of CMB fluctuations, therefore they get more attention. Density profiles, however, are not as dependent on the assumptions about the radial temperature profiles, so they are better probes of self-similarity in the cluster population and also receive considerable attention in the literature.

2.1.3 The structure of the chapter

In this chapter we study the observed density profiles of massive clusters across a wide range or redshifts, and characterize their self-similarity and deviations from self-similarity. Sec. 2.2 describes the cluster samples under study; Sec. 2.3 discusses model-independent

results for emission measure and the intrinsic scatter of gas density; Sec. 2.4 discusses 3D modelling of gas density for individual and stacked clusters; Sec. 2.5 discusses stacking clusters in bins by one of the global cluster parameters, and explores any differences between these bins; Finally, Sec. 4.6 summarizes our results.

Throughout the paper we adopt a flat Lambda-CDM cosmology with $H = 70\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_\Lambda = 0.73$, Ω_m=0.27, and Ω_b=0.0044.

2.2 The data

2.2.1 Cluster samples

To test the self-similarity of gas density profiles, we used a combination of 2 cluster samples. The first is the high-z subsample of the 400 square degree galaxy cluster survey abbreviated as 400d (Burenin et al., 2007). The other sample consists of the 80 most significant clusters detected in the 2500 deg^2 SPT-SZ field, and observed by Chandra in the Chandra X-ray Visionary Project (PI B. Benson) abbreviated here as SPT-XVP. The SPT-XVP sample consists of 80 clusters spanning the redshift range $0.3 < z < 1.2$ and the mass range $2 \times 10^{14} M_\odot < M_{500} < 20 \times 10^{14} M_\odot$. (Benson et al., 2014).

Both cluster samples spawn a similar redshift range and we have similar quality Chandra observations, with exposures typically sufficient to obtain ∼2000 X-ray source counts. From the joint redshift and mass distribution of clusters in both catalogs shown in Fig. 2.1, one can see that these catalogs have a good overlap in both parameters, but the typical mass of the 400d objects is somewhat smaller. Therefore, adding 400d clusters improves our statistics in the region where the catalogs overlap and additionally lets us explore clusters of slightly

lower masses.

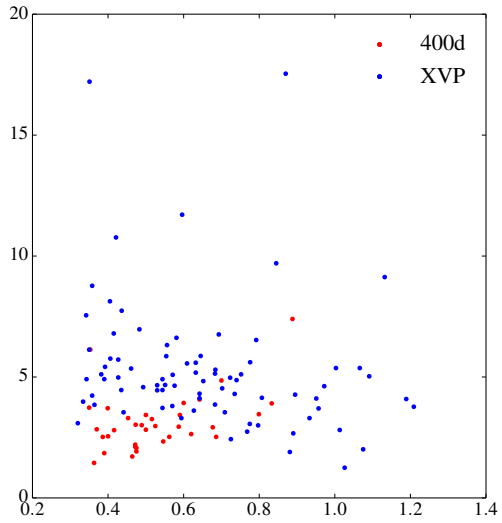

Figure 2.1: Distribution of redshift and mass for 400d and XVP clusters. These samples cover similar range of redshift. Including the 400d clusters allows us to extend the analysis to lower mass systems compared to XVP alone.

The selection criteria of X-ray luminosity and SZ signal-to-noise ratio are considerably different. Although both X-ray and SZ-selection are designed to produce quasi mass-limited samples, there is an ongoing discussion whether these different selection functions might be biased towards certain types of clusters. We partially addressed this question in Chapter 4. In this section we confirm that there is no difference in gas density profiles between these two samples as the first step of our analysis. After that we use the combined sample to increase statistical significance of our results.

We excluded several clusters from the XVP and 400d samples due to the presence of secondary clusters that was impossible to excise. Excluded clusters are: SPT-CLJ0330-5228, SPT-CLJ0551-5709 in XVP catalog; cl0521m2530, cl0141m3034, cl1312p3900 in 400d

catalog.

2.2.2 Determination of emission measure profiles

A number of steps need to be taken to proceed from the raw X-ray observations to emission measure profiles, namely:

1. Clean X-ray flares.
2. Apply Chandra calibrations.
3. Determine the background.
4. Eliminate point sources and clumpy substructures.
5. Determine the X-ray centroid.
6. Determine R_{500}.
7. Convert X-ray counts in each radial bin to the emission measure.

We rely on the analysis of McDonald et al. (2013) for X-ray data processing, i.e., the emission measure profiles we use here are identical to those of McDonald et al. (2013). We, however, briefly review the main stages of data analysis in this subsection.

Preparation of X-ray images

Steps 1 to 5 are performed by the data reduction pipeline which is identical to that used in McDonald et al. (2013) and Benson et al. (2014). Briefly, the raw observations are flare cleaned, corrected using the latest calibration corrections, the background is estimated from the off-source region, point sources are excluded using the wavelet decomposition technique.

Additional clumpy substructures were masked by hand and also excluded from the analysis. The center of the cluster was chosen by iteratively measuring T_X (or Y_X) in the $0.15-1\,R_{500}$ annulus and requiring that it satisfies $T_X - M$ (or $Y_X - M$) scaling relation. R_{500} is used for making radial bins for emission measure profiles and later for proper cosmological scaling of density profiles.

Surface brightness profiles

The surface brightness profile for each cluster is measured in a series of 32 annuli, covering the range$(0-3R_{500})$, with the outer radii for each annulus defined as:

$$r_i = 1.5\ \mathrm{R}_{500}\left(\frac{i}{20}\right)^{1.5}, \quad i = 1...32\ . \tag{2.4}$$

For each of these radial bins, X-ray counts are extracted in the energy range 0.7–2.0 keV, corrected for local exposure maps and background, and converted into emission measure through the normalization of the MEKAL model (see Benson et al. (2014) for details). We note that the conversion from count rate to emission measure is insensitive to temperature. (It changes by only a few percent for temperatures between 2 and 8 keV (Eckert et al., 2011)). The resulting emission measure (in cm^{-5}) is essentially the value of line of sight integral

$$EM = \int n_e n_p dl, \tag{2.5}$$

where n_e and n_p are the electron and proton densities, respectively.

Surface brightness for individual clusters in the bins beyond $2R_{500}$ is indistinguishable from both the background determined from the off-source regions of the same field, and from Chandra blank-sky backgrounds. However, we still retain the data in these radial bins, because by using these data while fitting the emission model, the residual background

can be constrained. Surface brightness beyond $3R_{500}$ includes increased unresolved X-ray background due to the off-axis PSF degradation therefore does not improve the background constraints.

The same procedure was used for determining the emission measure profile of the 400d clusters, but the radial binning scheme was different with much narrower bins. We convert profiles of the 400d clusters (which are identical to the ones used in Vikhlinin et al. (2006)) to the same binning scheme (2.4) by taking the appropriate weighted average of surface brightness values in the old bins that overlap with each of the new bins. This procedure virtually does not affect the uncertainties of surface brightness in the new bins (which are properly propagated from the old uncertainties), because the original 400d bins were much narrower.

Residual background subtraction

Elimination of background by subtracting black-sky Chandra fields is not accurate enough for modelling the background around or beyond R_{500} – the extracted emission measure profiles tend to a nonzero constant at large ($\sim 3R_{500}$) radii. Measuring this constant is non-trivial because of large uncertainties and the residual cluster emission at these radii. We solve this problem by including the background as an additional component in our 3-D emission model (see Sec. 2.4.2) and fitting that emission model to each cluster individually. We then apply this background correction to each individual cluster's emission measure profile, which is necessary for the analysis of intrinsic scatter in the universal emission measure profiles (Sec. 2.3.2). If one wishes to dispense with 3-D modelling, however, fitting of the power law + constant to surface brightness from $0.8R_{500}$ to $3R_{500}$ gives similar background

values.

The fitted background changes slightly depending on the model of cluster emission that we use. We include this variability in the uncertainty of the background (which can also be thought of as the systematic uncertainty of the projected emission profiles.)

2.3 Model independent comparison of emission measure profiles.

2.3.1 Scaled emission measure profiles

In Sec. 2.1 we show that the total matter density in a cluster should be approximately described by a universal function of r/R_{vir} and that we can use the overdensity radius $R_\Delta,\ \Delta = 500$ instead of R_{vir} as a scaling factor for r. We have not discussed, however, how the overall normalization of this function changes with redshift. The latter follows from the definition of R_{500} as the radius of a sphere with a fixed mean overdensity with respect to the critical density of the universe. Indeed

$$\begin{gathered}\Delta\frac{4}{3}\pi\rho_c R_\Delta^3 = M_\Delta = \int_0^{R_\Delta}\rho(r/R_\Delta)4\pi r^2 dr = R_\Delta^3\int_0^1\rho(x)4\pi x^2 dx \\ \Delta\frac{4}{3}\pi\rho_c = \int_0^1\rho(x)4\pi x^2 dx,\end{gathered} \tag{2.6}$$

From the last equation, it is evident that the density of matter at a fixed $x = r/R_{500}$ is proportional to ρ_c.

The ICM gas follows the distribution of dark matter, at least outside cluster cores. Therefore we expect the same scaling laws to apply for electron density n_e

$$n_e(r) = n_c f(r/R_{500}), \quad n_c = \frac{\Omega_b}{\Omega_m}\frac{\rho_c}{2\mu m_p}, \tag{2.7}$$

where $f(x)$ is the universal radial gas distribution, n_c is the concentration of electrons in the critical-density Universe, Ω_b and Ω_m are baryon and total matter fractions m_p is the proton mass and $\mu \simeq 0.6$ is the mean molecular mass for a gas of primordial composition. The normalization factor n_c is defined this way so that $f(x) = \rho_{tot}/\rho_c$ and therefore can be directly compared to overdensity of dark matter halos in cosmological simulations.

Emission measure EM is the line of sight integral of $n_e n_p$. As a function of projected distance R from cluster's center it can be expressed as

$$\begin{aligned} EM = \int n_e n_p dl = 2 \int_R^\infty \frac{n_e(r) n_p(r) r dr}{\sqrt{r^2 - R^2}} \\ = R_{500}\, n_c^2\, 2 \int_{R/R_{500}}^\infty \frac{f^2(x) x dx}{\sqrt{x^2 - (R/R_{500})^2}} = R_{500}\, n_c^2\, S_X(R/R_{500}), \end{aligned} \tag{2.8}$$

where $S_X(x)$ is the universal radial emission measure profile.

The dimensionless universal radial emission measure profile (or scaled emission measure profile) can be obtained from the measured emission measure $EM = \int n_e n_p dl$ by

$$S_X = n_c(z)^{-2}\, R_{500}^{-1}\, EM. \tag{2.9}$$

As already discussed in Sec. 2.1, we use $R_{500}(Y_X)$ as an estimate of R_{500} in eq. (2.9).

Fig. 2.2A, shows the original surface brightness profiles in physical units: radius in Mpc on x-axis and surface brightness in cm^{-5} on y-axis. One can see a significant variation of emission measure at the same physical radius, with hotter clusters (red) having a higher normalization than the colder cluster (blue). Fig. 2.2B, shows surface brightness profiles with radius scaled to $R_{500}(T_X)$ and surface brightness scaled according to Eq. (2.9). It is easy to see that scatter decreases indeed. In Fig. 2.2C, surface brightness profiles are scaled to $R_{500}(Y_X)$ and show a further decrease in scatter, consistent with the belief that Y_X gives the most accurate estimate of cluster mass. We will discuss the observed scatter in scaled emission measure profiles more quantitatively in the next Section.

In all panels cluster profiles are colored according to their temperature. All left panels of Fig. 2.2 may give an impression that the residual surface brightness at radii beyond R_{500} is nonzero. In reality, the residual S_x is scattered around 0, but only points with positive value are shown on the log-log plot.

2.3.2 The intrinsic scatter of emission measure profiles

The observed scatter of emission measure S_X in each radial bin sums up the intrinsic scatter of cluster emissivity for the entire cluster population and the statistical noise of measuring the flux for each individual cluster:

$$\sigma^2_{obs}(S_X) = \sigma^2_{ex}(S_X) + \frac{1}{N}\left(\sum_i \left(\frac{\Delta S_{X,i}}{S_{X,i}}\right)^2 + \left(\frac{\Delta B_i}{B_i}\right)^2\right), \tag{2.10}$$

where $\sigma_{ex}(S_X)$ is the excess scatter in emission measure at given radius, not accounted for by Poisson noise and background uncertainty (the origin on this scatter is discussed in the next paragraph); $\Delta S_{X,i}/S_{X,i}$ is the shot noise in the given radial bin for i-th cluster (square root of counts); and $\Delta B_i/B_i$ is the uncertainty of the background for the given cluster which can be determined by bootstrapping methods (see Sec. 2.3.5). We call the average combined uncertainties in $S_{X,i}$ and B_i (second term in eq. 2.10) *statistical* scatter.

The emission measure of the i-th cluster is the product of a universal shape of cluster profiles

$$S_{X0} = \int \frac{n_e n_p}{n_c^2} \frac{dl}{R^T_{500}} \tag{2.11}$$

and the inverse ratio of R^T_{500} - true and R^M_{500} - measured R_{500}

$$S_{X,i} = \frac{R^M_{500,i}}{R^T_{500}} S_{X0}. \tag{2.12}$$

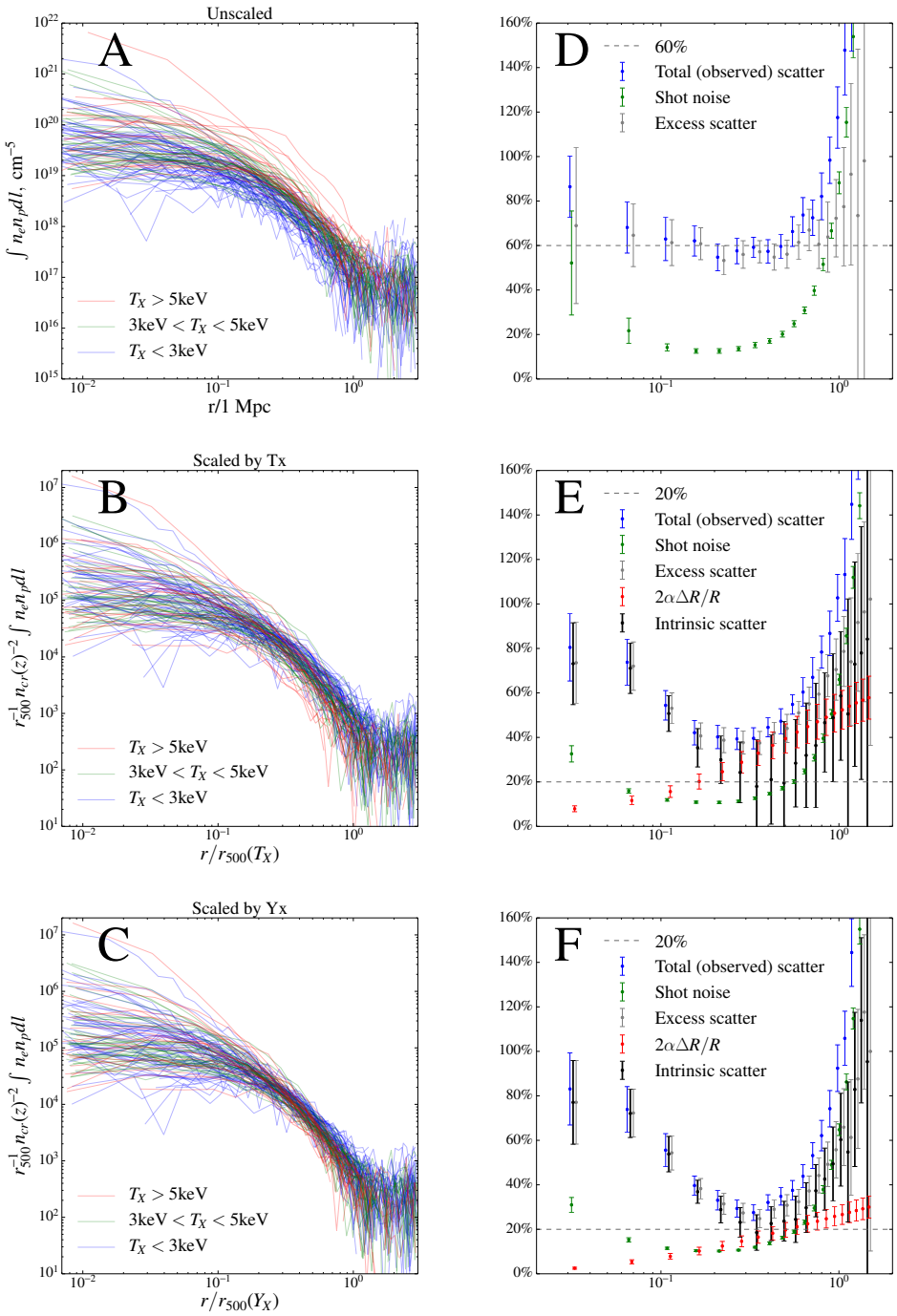

Figure 2.2: Scaling emission measure profiles by different characteristic radii and sources of scatter in emission measure profiles. A - no scaling. The scatter in S_X is high. B - self similar scaling performed with $R_{500}(T_X)$. The scatter decreases. C - self similar scaling performed with $R_{500}(Y_X)$. The scatter decreases further. D, E, F - show different contributions to scatter for each of the applied scalings. Plots D and E show that the intrinsic scatter in S_X is $\sim 20\%$. Dashed lines at 20% and 60% are added to guide the eye.

To propagate uncertainties in eq.(2.12) we need to take into account that $S_{X,i}$ in the left hand side represents the emission measure as a function of *measured* distance $x^M = r/R^M_{500}$ from the center in units of R_{500}, and S_{X0} in the right hand side represents the universal emission measure as a function of *true* distance $x^T = r/R^T_{500}$ from the center. Taking logarithm of both sides of eq. (2.12), obtain

$$\begin{aligned} \ln S_{X,i}(x^M) &= \ln R^M_{500}/R^T_{500} + \ln S_{X0}(x^T) \\ &= \ln R^M_{500}/R^T_{500} + \ln S_{X0}(x^T) + \frac{\partial \ln S_{X0}}{\partial \ln r}\frac{x^M - x^T}{x^T} \\ &= \ln S_{X0}(x^T) + \left(1 + \frac{\partial \ln S_{X0}}{\partial \ln r}\right)\frac{\Delta x}{x}. \end{aligned} \tag{2.13}$$

It follows from eq. (2.8), at large radii where the logarithmic slope of ρ_g doesn't change too fast (practically at $r > 0.5R_{500}$), that the logarithmic slope of S_X is related to the logarithmic slope of ρ_g by

$$\frac{\partial \ln S_{X0}}{\partial \ln r} \approx 2\frac{\partial \ln \rho_g}{\partial \ln r} + 1 = -2\alpha + 1. \tag{2.14}$$

Substituting (2.13) to (2.14) and taking into account the intrinsic scatter $\sigma_{intr}(S_{X0})$ in the universal emission measure profile S_{X0}, we obtain the expression for the excess scatter

$$\sigma_{ex}(S_{X0}) = \sigma^2_{intr}(S_X) + \left(2\alpha\frac{\Delta R_{500}}{R_{500}}\right)^2. \tag{2.15}$$

Summarizing equations (2.10) - (2.15), we get the following formula for intrinsic scatter in S_X

$$\sigma^2_{intr}(S_X) = \sigma^2_{obs}(S_X) - \frac{1}{N}\left(\sum_i \left(\frac{\Delta S_{X,i}}{S_{X,i}}\right)^2 + \left(\frac{\Delta B_i}{B_i}\right)^2\right) - \left(2\alpha\frac{\Delta R_{500}}{R_{500}}\right)^2. \tag{2.16}$$

Therefore, the scatter in the measured emission measure profiles $\sigma_{obs}(S_X)$ is explained by the intrinsic scatter $\sigma^2_{S_{X0}}$ in the universal profile S_{X0}, the relative error in measuring R_{500}, and the logarithmic slope of the density profile.

The uncertainty of R_{500} in turn depends on the scaling relation used to obtain it. $M_{T_X} \propto T_X^{3/2}$ (3/2 is the theoretical slope, but within 1σ of observed) and $R_{T_X} \propto M_{T_X}^{1/3}$, So for $R_{500}(T_X)$ we obtain

$$\left(\frac{\Delta R_{500}(T_X)}{R_{500}(T_X)}\right)^2 = \left(\frac{1}{2}\frac{\Delta T_X}{T_X}\right)^2 + \left(\frac{1}{3}\sigma_{M-T_X}\right)^2, \tag{2.17}$$

where $\Delta T_X/T_X = 19 \pm 7\%$ is the error in measuring temperature, and $\sigma^2_{M-T_X} = 17 \pm 5\%$ is the intrinsic scatter in the scaling relation (Kravtsov et al., 2006).

Analogously, $M_{Y_X} \propto T_X^{3/5}$, so for $R_{500}(Y_X)$ we get

$$\left(\frac{\Delta R_{500}(Y_X)}{R_{500}(Y_X)}\right)^2 = \left(\frac{1}{5}\frac{\Delta Y_X}{Y_X}\right)^2 + \left(\frac{1}{3}\sigma_{M-Y_X}\right)^2, \tag{2.18}$$

where $\Delta Y_X/Y_X = 26 \pm 10\%$ is the error in measuring Y_X, and $\sigma^2_{M-Y_X} = 6 \pm 1\%$ is the intrinsic scatter in the scaling relation (Kravtsov et al., 2006).

Strictly speaking, equations (2.10) - (2.18) only work for Gaussian distributions with σ defined as standard deviations of those distribution. We avoid using standard deviation for estimates of scatter σ due to heavy-tailedness of our distributions. Instead, we use its robust counterpart based on *Median Absolute Deviation*

$$\begin{gathered}\sigma \approx 1.4826 \ \text{MAD}, \\ \text{MAD} = \text{median}_i \left(\ |X_i - \text{median}_j(X_j)| \ \right),\end{gathered} \tag{2.19}$$

We assume that error propagation methods still work in the moderate deviations from Gaussianity with σ_{MAD} as an estimator of standard deviation.

Substituting uncertainties of Y_X, T_X and scatter into scaling relations to eq. (2.17), eq. (2.18), we get

$$\begin{aligned}\frac{\Delta R_{500}(T_X)}{R_{500}(T_X)} &= 12 \pm 2\% \\ \frac{\Delta R_{500}(Y_X)}{R_{500}(Y_X)} &= 6 \pm 1\%.\end{aligned} \tag{2.20}$$

We will use these uncertainties in R_{500} in eq. 2.16 for the computation of the intrinsic scatter.

Fig. 2.2D shows the observed, statistical and excess scatter after scaling the emission measure profiles to 1Mpc. The excess scatter stays at the level of $\sim 50\%$ for a wide range of radii, approximately from 100kpc to 800 kpc. The excess scatter seems to increase beyond 800kpc, which we associate with the fact that cluster emission and background are not easily separable at these radii (and for some clusters background $\gg$ cluster emission), so the simple uncertainty framework developed above breaks down and does not give a reliable estimate of the excess scatter.

Fig. 2.2E, F show observed, statistical and R_{500}-related scatters and computed from them intrinsic scatter as a function of radius. We note that for both scalings intrinsic scatter stays at the level of $\sim 20\%$ for a radial range $0.3 - 0.7R_{500}$. At greater radii intrinsic scatter increases due to the effects of background modelling. We see that the large observed scatter of 40% in emission profiles scaled to $R_{500}(T_X)$ compared to only 25% in emission profiles scaled to $R_{500}(Y_X)$ is completely explained by greater uncertainty in $R_{500}(T_X)$.

We repeated calculations of intrinsic scatter separately for the XVP and 400d catalogs, and found them to be within uncertainties of intrinsic scatter for the combined catalog.

2.3.3 Cluster stacking

In the rest of the paper we will use stacked emission measure profiles to eliminate the uncertainties associated with Poisson noise (especially for large radii) and uncertainties in R_{500}. Stacking not only improves the signal-to-noise ratio, but also makes the approximation that the clusters are spherical more valid, because we average over all ellipticities and

orientations that exist in our dataset.

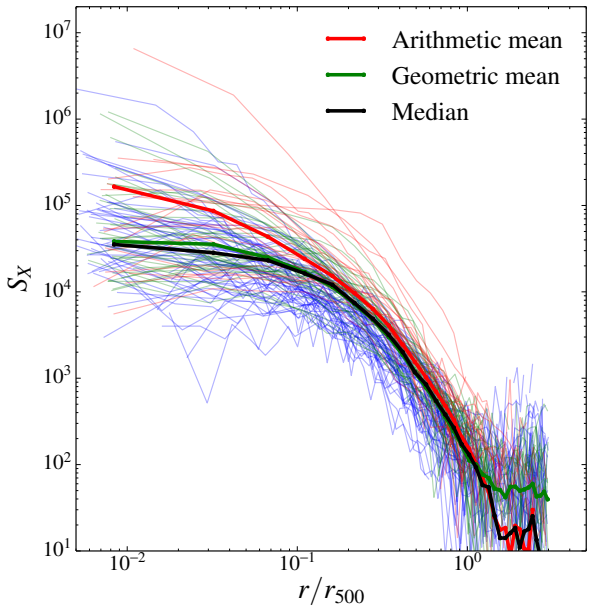

Figure 2.3: Different averaging methods. See Sec. 2.3.4 for discussion.

2.3.4 Averaging profiles in stacks

We considered 3 methods of averaging surface brightness profiles in stacks:

- Arithmetic mean
- Geometric mean
- Median

The simplest method of profile averaging is the arithmetic mean of individual profiles. However, there are extreme cool core clusters in our sample that are ∼100 times brighter

in their centers than average. They dominate in the arithmetic mean, making the average profile dominated by a few cool core clusters. We do not aim at an accurate description of cluster at the cores, but even at moderate radii (0.3-1 R_{500}), the difference between the brightest and the dimmest clusters is an order of magnitude. Therefore, simple arithmetic average is always biased towards brighter clusters, and is an inaccurate representation of a typical cluster's surface brightness profile.

Another reason why the arithmetic mean is not optimal is that the distribution of surface brightnesses at a given radius is closer to lognormal than to normal (especially at radii where the background can be neglected) This suggests that the geometric mean:

$$\langle SB \rangle_{geom} = \exp(\langle log(SB) \rangle) \tag{2.21}$$

should produce a more representative averaged profile.

Unfortunately, a geometric mean has another undesirable quality: it is undefined if there is even a single negative value of surface brightness. Such values often occur in the outskirts where the cluster emission is lower than the uncertainty of the background.

The averaging method which is free from the previously discussed drawbacks is the median. The median is less sensitive to the actual distribution of the emission at a given radius, coincides with geometric mean for lognormal distribution and arithmetic mean for normal distribution, insensitive to the outliers. This statistic works well for the entire range of radii. Figure 2.3 shows the comparison of the three averaging methods. The median accurately traces the "typical" surface brightness profile for the entire range of radii, while the arithmetic mean "overestimates" the density at small radii, and the geometric mean is undefined at large radii, according to the discussion above. All three methods nearly

agree for moderate radii (0.3 - 0.8 R_{500}) which confirms that the median is a reasonable representation of a typical cluster profile.

One might argue that the median is a suboptimal way of averaging in the cluster cores, because there are fewer CC systems compared to NCC systems, so the median would take on an NCC profile. We, however, argue that there is a continuous (not bimodal) distribution of surface brightness at small radii, so the median produces a correct result. Also, as the reader can see in Fig 2.3, the median almost coincides with the geometric mean in cluster cores.

The distribution of errors on mass, and consequently on R_{500}, is approximately lognormal (they are obtained from scaling relations). In other words, one is as likely to overestimate R_{500} by, say, 1% as to underestimate it by 1%. If cluster profiles are self-similar (with no intrinsic scatter) with the outer slope of emission measure γ, the aforementioned errors on R_{500} would produce the $\pm\gamma$% errors on scaled emission measure S_X. The mean of $1 \pm \gamma/100$ is greater than 1, which means that the mean is biased. The geometric mean and the median handle lognormal scatter in R_{500} better.

2.3.5 The use of bootstrapping for uncertainty estimation

Although the cluster properties show a large intrinsic scatter, the large sample of clusters lets us narrow down the uncertainties on the *average* properties of these clusters.

We use bootstrapping to find the median profile and associated uncertainties in the following way.

1. We start with N background-corrected profiles of individual clusters (see Sec. 2.2.2)
2. We draw N clusters from the set with replacement. (This step simulates the variability

in the composition of the catalog)

3. For each cluster we generate the surface brightness in each radial bin by adding Gaussian Noise (with sigma equal to uncertainty of surface brightness) to its observed surface brightness. (This steps simulates the variability in shot noise.)

4. We find the median of all values obtained in step 3. This is our resampled surface brightness.

5. We fit the model of cluster emission that includes the background component to the data. For more details on the models and fitting procedure see Sec. 2.4.2. Here we only use the background obtained in the fitting procedure which subtract from all radial bins.

6. We repeat steps 2-5 M times ($M = 100$ is a typical number of repetitions)

7. We use the median of all resampled emission measures as the value of emission measure of the average cluster. We use 16th and 84th percentiles as 1-sigma errors.

2.3.6 Results for stacked emission profile and X-ray emission beyond R_{500}

In Fig. 2.4A we show the result of this procedure applied to all clusters in the combined XVP+400d catalog. There are 2 types of uncertainties that can be associated with the emission measure in each radial bin: statistical-only uncertainty (black error bars) and full bootstrapped uncertainty (green error bars). The blue line is the model fit. Statistical-only uncertainties quantify Poisson noise only and show how well we know the median emission measure for the *given* sample of clusters, assuming that no additional uncertainties are

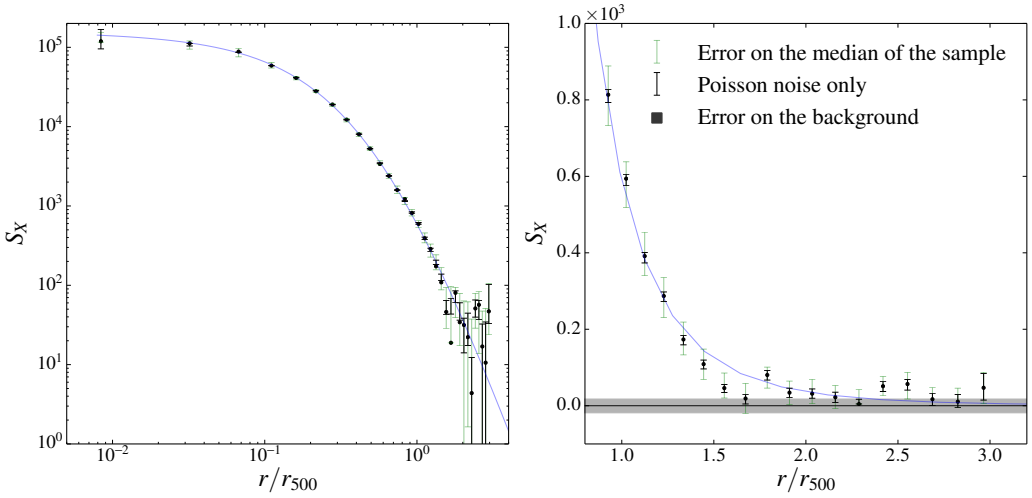

Figure 2.4: Emission measure for a stack of all clusters. This figure shows model independent results for the entire stack of clusters. Black error bars represent shot-noise-only uncertainties on the average emission measure profiles, while green error bars also include all possible sources of uncertainties. The blue line represents model fit to the observed median emission measure profile.

introduced during the data processing steps. Full bootstrapped uncertainties provide the expected bounds on the median emission measure of *any* sample of ∼115 clusters that could be obtained in a similar survey, and take into account additional uncertainty associated with all stages of data processing. As follows from Sec. 2.3.2, the main source of uncertainty is R_{500} which enters the formula for scaled emission measure. If these full uncertainties were not computed, one could imagine that all of the observed signal beyond R_{500} is the result of a severe underestimation of R_{500} for a few clusters, i.e. what we see is the emission inside R_{500} that we mistakenly attribute to the outer region. Having computed these full uncertainties, we can claim that the emission beyond R_{500} is definitely present in our data, and that we would be able to see it with any cluster sample of similar size and similar mass and redshift ranges.

Observing the outskirts of galaxy clusters gives us an opportunity to understand the transition between gravitationally-bound gas in the state of hydrostatic equilibrium and clumpy infalling material. It is also important for the calibration of X-ray mass measurements (Eckert et al., 2011) and especially SZ-mass measurements for which the ICM behavior at large radius has stronger effect. High significance detection beyond R_{500} is also important because we can measure logarithmic slope α at these radii. However, the only X-ray telescope which is able to routinely detect X-ray emission up to the virial radius for distant clusters is *Suzaku* which has low angular resolution.

Here we report a detection of X-ray emission up to 1.4 R_{500}. The right panel of Fig. 2.4 zooms on the cluster outskirts. One can see 4 bins beyond R_{500} with better than 3σ detection of X-ray emission.

Our detection of X-ray emission almost up to the virial radius ($R_{vir} \approx 1.6R_{500}$) with *Chandra* for a sample of distant clusters is the first of its kind and can provide important cross-calibration with *Suzaku* and cluster simulations.

2.4 Model-dependent 3-D gas density profiles

In the previous section we discussed model-independent methods and results based directly on surface brightness profiles. In this section we look at gas density which needs to be obtained by deprojection of the surface brightness profiles. There are model-independent methods of deprojection, but they cannot be applied to our observations which on average have ~ 2000 counts. Therefore we have to resort to model-dependent methods.

2.4.1 Modelling 3-D gas density profiles

While a wide range of models have been used in the literature, most of them can be described in terms of either a generalized NFW model (Nagai et al., 2007)

$$\rho_r(r) = \rho_0 \frac{(r/r_c)^{-\alpha}}{(1+(r/r_c)^{\gamma})^{\beta/\gamma}} \tag{2.22}$$

or generalized beta model (Vikhlinin et al., 2009)

$$\rho_r(r) = n_0^2 \frac{(r/r_c)^{-\alpha/2}}{(1+r^2/r_c^2)^{3\beta/2-\alpha/4}} \frac{1}{(1+r^3/r_s^3)^{\epsilon/6}}, \tag{2.23}$$

by fixing some of the exponents to the predetermined values (Merritt et al., 2006).

Equations 2.22 and 2.23 hold some similarity. In our efforts to optimize the functional form of the fitted profile we found a parametrization that sheds some light on why these types of functions may be useful in fitting radial profiles of gas density, pressure, dark matter etc.

The variety of analytic functions used to fit radial profiles, which at first sight seem a little arbitrary (see Chapter 1, and also Silva et al. (2013); Merritt et al. (2006) for more details), looks more reasonable when we think of radial profiles in terms of their logarithmic derivative alpha

$$\alpha = -\frac{d\ln\rho}{d\ln r}. \tag{2.24}$$

Instead of discussing radial profiles in their original coordinates ρ, r, we will discuss their meaning in the coordinates (α, ξ), where $\xi = \ln(r/r_0)$, and r_0 is a characteristic size of the object.

One of the advantages of such a parametrization is that it maps astronomically huge dynamic ranges of densities and radii to a much more convenient range of $\sim$ [-4, 1] for ξ (i.e. we typically have reliable measurements of density or pressure or temperature starting

from $\sim 10^{-2}$ of the characteristic radius up to a few characteristic radii) and $\sim [0, 4]$ for α. The other advantage is that many widely used profiles are expressed as much simpler functions than in the original coordinates, and the choice of these functions becomes more transparent.

The two fundamental requirements for α as a function of ξ are

$$\begin{aligned} \alpha(\xi) &\to const > 0, \qquad \text{as } \xi \to -\infty \\ \alpha(\xi) &\geq 3, \qquad \text{as } \xi \to \infty. \end{aligned} \tag{2.25}$$

The first condition demands that alpha stays defined as $r \to 0$ and that the density not decrease towards the center of the halo. The second condition demands that the total mass of the halo enclosed within radius r doesn't diverge faster than $\ln r$.

There are only a few simple analytic functions that satisfy both requirements. The simplest one is, of course, a constant, which defines a power law density profile similar to those used in the early structure formation models

$$\alpha(\xi) = A \quad \Rightarrow \quad \rho \sim r^{-A}. \tag{2.26}$$

However, in many real systems α changes with radius (or ξ). Typically, density is saturated in the center, which means that $\alpha(r = 0) = 0$ and increases outwards. This behavior is described by the exponential function which generates the Einasto profile

$$\alpha(\xi) = A \exp(\gamma\xi) = Ar^{\gamma} \quad \Rightarrow \quad \rho \sim \exp(Ar^{-\gamma}). \tag{2.27}$$

The main feature of the Einasto model is a continuously increasing α at large radii leading to an increasing steepness of $\rho(r)$. This is not necessarily confirmed in observations and numerical models. After all, both the NFW and beta-models have a fixed logarithmic

slope at large radii. Therefore, it is also important to consider an alternative class of models where α transitions to a constant as $r \to \infty$.

The simplest function that approaches two different constants at $\pm\infty$ and makes a smooth transition between them is a hyperbolic tangent. The following functional form

$$\alpha(\xi) = A + B \tanh \eta(\xi - \xi_0) \tag{2.28}$$

parametrizes the position of the tangent on the ξ axis and its horizontal and vertical scales. A simple calculation shows that the corresponding density profile reads

$$\rho \sim r^{A'}(1 + (r/r_0)^{2\eta})^{-B'/2\eta}, \text{ where } \mathrm{A'} = \mathrm{A} - \mathrm{B}, \quad \mathrm{B'} = \mathrm{A} + \mathrm{B}, \quad \mathrm{r_0} = \mathrm{e}^{\xi_0} \tag{2.29}$$

This observation shows that the NFW profile, the generalized NFW, the beta-model, and the cuspy beta model (which includes the first factor in Eq.2.23) are all accomodated by Eq.2.28. As expected, $A' = \lim_{\xi\to-\infty} \alpha(\xi)$ controls the inner slope, $B' = \lim_{\xi\to\infty} \alpha(\xi)$ controls the outer slope, and η (which corresponds to γ the parameters of the GNFW and the generalized beta-models) corresponds to the width of the transition region from one logarithmic slope to the other. For example, the beta model is described by Eq. 2.28 when we choose $A' = 0$, $B' = 3$, $\eta = 1$. The NFW model is described by Eq. 2.28 when we choose $A' = 1$, $B' = 3$, $\eta = 0.5$

One could use more terms in the equation for $\alpha = \alpha(\xi)$ to fit the observed logarithmic slope. The sum of several terms in (α, ξ)-space corresponds to the multiplication of the corresponding densities in (ρ, r)-space. Consequently, the generalized beta-model in the form (5) is described by the following function $\alpha(\xi)$

$$\alpha = A + B \tanh(\xi - \xi_c) + C \tanh \frac{3}{2}(\xi - \xi_s). \tag{2.30}$$

The natural question that arises at this point is what other combination of functions could be used to constrain the α-profile. One could use a "tanh series" as the generalized beta-model suggests, but it may be suboptimal because of the high degeneracy between the functions from this family.

Since in any observation we only have data constraining α on a certain interval $[\xi_{min}, \xi_{max}]$, one of the options is to use a set of functions that gives α considerable freedom over this interval, with a smooth extrapolation outside of this interval which satisfies the two physical condition of Eq. 2.25.

It is now easy to understand the most important difference between GNFW and GB models: GB model is able to fit α-profile with two "steps", one of which has width 1 in log coordinates and the other has width 2. The GNFW model can only fit a single "step", but it is capable of making this step sharper. This results in changing the model parameters as a function of maximum fitted radius when fitting to profiles with continuously changing logarithmic slope. This is of no concern though, because the value of α for given ξ stays approximately constant in spite of changing model parameters. A GNFW model has been widely used in the past, often with some of the parameters fixed. (See Merritt et al. (2006) for a review.)

Both features prove to be useful for some profiles, but combining them in a single model further increases the number of parameters and leads to overfitting for the dataset that we consider in this paper.

All individual clusters and all cluster stacks were adequately fitted with both a generalized beta model (GB) (2.23), and generalized NFW model (GNFW) (2.22). We also use an additional background component that we add after numerical deprojection.

2.4.2 Fitting and bootstrapping individual clusters and stacked cluster model

We follow the algorithm described in Sec. 2.3.5 to obtain the gas density model of an average cluster in the given stack. In step 5 of the algorithm, we fit the numerically projected generalized beta model or generalized NFW model by nonlinear least squares. In step 6 we have 100 different fitting curves for surface brightness, gas density and logarithmic slope α. For each radius we use the median value from these curves as the true value, 16th and 84th quantile as 1-sigma errors, 2nd and 98th quantile as 2-sigma errors.

Figure 2.5 shows the results of this procedure for the combined catalog. The two top panels show the scaled emission measure for the whole range of radii and zoomed in on the cluster outskirts. These panels contain both the confidence interval of the fitted models and the data used for fitting. The bottom two panels show confidence intervals of gas density and the α-parameter. All fitted quantities are shown with both 1-σ and 2-σ confidence intervals.

Note, that we analyze the correctness of combining the 400d and XVP cluster samples in Sec. 2.5.2.

The procedure of fitting individual clusters is identical to the one described above except that the stack now consists of a single cluster. Therefore, step 2 of the bootstrapping algorithm does nothing, but step 3 is still active and generates different realization of Poisson noise.

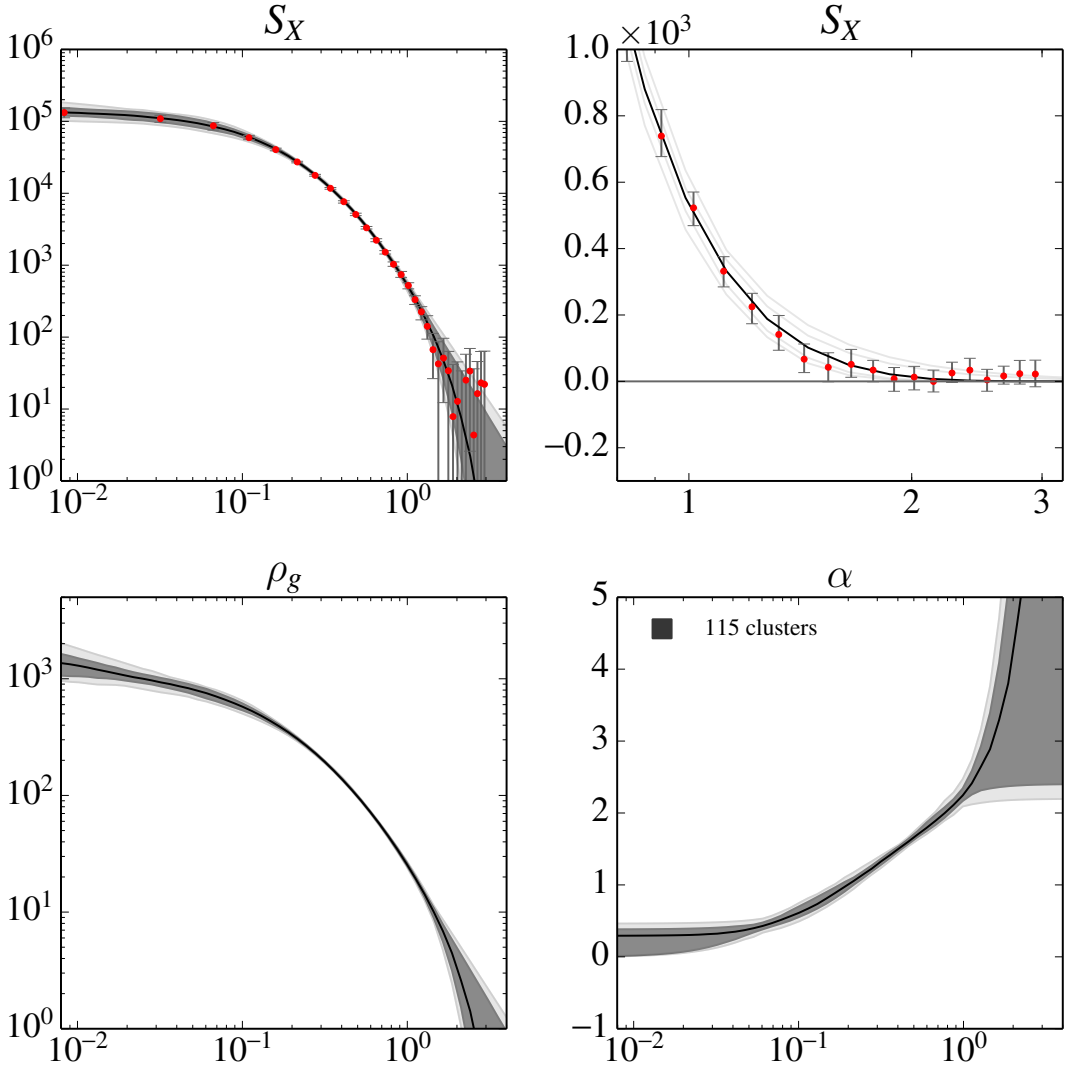

Figure 2.5: Full catalog fitted with generalized beta model. This figure shows the results of fitting generalized beta model to the full stack of 400d and XVP clusters. The top left panel shows the scaled emission measure profile (red dots with error bars) and the confidence interval of the model-derived emission measure profile. 1σ confidence interval is shown in dark gray and 2σ confidence interval is shown in lighter gray. The top right panel shows the same emission measure profile, but zoomed in on the cluster's outskirts and with linear scale on y-axis. The bottom left panel shows the model-derived gas density normalized by the critical density ρ_c. Finally, the bottom right panel shows the model-derived logarithmic slope α. The plot shows that the emission measure profile averaged over the entire stack is described and constrained well by generalized beta model. The α plot shows that the fit is well constrained up to R_{500}. Beyond that radius we still get good constraints on ρ_g and S_X, but uncertainty in α increases dramatically.

2.4.3 Intrinsic scatter in ρ_g

Given best-fit density profiles ρ_g with uncertainties we can revisit the question of intrinsic scatter of scaled gas density around the universal profile previously discussed in Sec. 2.3.2. Following the methodology outlined there we first move from the observed scatter to observed minus statistical scatter, and then subtract scatter associated with uncertainty of R_{500}. The only difference with eq. 2.15 is a coefficient α instead of 2α.

In Fig. 2.6 we show the scatter at different stages of this process. This analysis gives $8 \pm 4\%$ as intrinsic scatter of ρ_g. This is in excellent agreement with $20 \pm 10\%$ scatter in S_X.

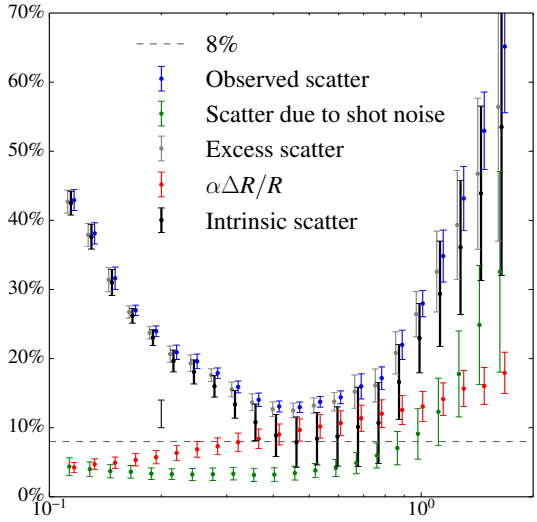

Figure 2.6: Intrinsic scatter in gas density. This figure shows all sources of scatter that contribute to the observed scatter in ρ_g

2.5 Deviations from self-similarity

In this section we study stacked profiles for various subsamples. We compare the average profiles of the XVP and 400d samples. We also check how the average profile varies with cluster mass, redshift, "cool coredness" and dynamical state. The summary of all the partitionings that we study is presented in Table 2.1

2.5.1 Methodology

In order to test for differences between subsamples, we first compute the models for $\rho_g(\xi)$ and $\alpha(\xi)$ with uncertainties for each subsample. We provide a visual comparison for these models. To provide a qualitative measure of the difference, we maximize the significance of the difference between the 2 functions based on their median and scatter σ.

$$\zeta^2 = \max_{0.3<\xi<1} \frac{\sigma^2(f_1(\xi)) + \sigma^2(f_2(\xi))}{|\langle f_1(\xi)\rangle - \langle f_2(\xi)\rangle|}, \tag{2.31}$$

where σ represents MAD-based scatter (eq. 2.19) and $\langle\rangle$ - median.

We only do the maximization over the region $\xi \in [0.3, 1]$ because we are interested in gravitational self-similarity which does not manifest in the cluster centers. We stress that ζ-parameter is not a precise statistical test that gives a significance of deviation, but rather an empirical method of comparing two profiles. Higher values of ζ should correspond, however, to lower probability that two radial profiles are just imperfect measurements of the same universal profiles. We apply this comparison technique for S_X, ρ_g and α profiles for all tested partitionings of our cluster sample. The results are summarized in Table 2.2.

Table 2.1: Summary of all the partitionings of the combined 400d+XVP catalog presented in this study.

Partitioning	Description
XVP/400d	2 bins: XVP and 400d catalogs
XVP/400d (2)	2 bins: overlapping parts of XVP and 400d catalogs
M	3 bins: $M_1 < 3 \times 10^{14} M_\odot < M_2 < 5 \times 10^{14} M_\odot < M_3$
z	3 bins: $z_1 < 0.5 < z_2 < 0.75 < z_3$
$\mu = M/M_{\rm NL}$	3 bins: $\mu_1 < 500 < \mu_2 < 1500 < \mu_3$
K_0	2 bins: $K_{0\,1} < 30\,[\mathrm{keV\,cm^2}] < \mathrm{K}_{0\,2}$
$\mathrm{A_{phot}}$	3 bins: $A_{phot\,1} < 0.15 < A_{phot\,2} < 0.6 < A_{phot\,3}$
NCF, $\mathrm{A_{phot}}$	3 bins: $A_{phot\,1} < 0.15 < A_{phot\,2} < 0.6 < A_{phot\,3}$ AND $K_0 > 30\,[\mathrm{keV\,cm^2}]$

Table 2.2: Deviations of radial profiles in various bins. The table shows the maximum deviations of α, ρ_g and S_X radial profiles between the bins for various partitioning ofthe entire cluster catalog. r is the radius at which profiles differ the most, ζ is the "significance" of deviation (see eq. 2.31, this is not the true statistical significance), $\Delta\alpha$ is the absolut value of the maximum difference between α-profiles, $\Delta\rho_g$ and ΔS_X is the maximum difference in percent between the bins. For the definitions of the partitionings see text and Table 2.1

Partitioning	α			ρ_g			S_X		
	r	ζ	$\Delta\alpha$	r	ζ	$\Delta\rho_g$	r	ζ	ΔS_X
XVP/400d	0.41	1.84	0.10	0.68	1.81	5%	0.60	1.61	11%
XVP/400d (2)	0.32	2.64	0.24	0.46	3.14	11%	0.46	3.03	22%
M	0.52	2.98	0.18	0.32	2.85	9%	0.32	1.72	12%
z	0.32	3.22	0.30	0.60	1.48	7%	0.46	1.36	13%
$M/M_{\rm NL}$	0.52	2.21	0.24	0.32	2.38	12%	0.32	2.26	20%
K_0	0.32	3.87	0.19	0.52	2.92	10%	0.46	2.93	18%
$\mathrm{A_{phot}}$	0.32	3.50	0.39	0.60	2.79	14%	0.52	2.73	24%
NCF, $\mathrm{A_{phot}}$	0.32	2.65	0.31	0.41	2.48	14%	0.32	2.24	22%

2.5.2 Comparison of 400d and XVP Catalogs

The combination of the XVP and 400d catalogs increases the number of objects in our analysis and can improve the statistical significance of our results. However, before combining the clusters from these two samples, we need to check that the differences between X-ray and SZ-selections did not result in clusters that are significantly different from one another. To assess this we do two tests: 1) we compare the stacked profiles of the entire XVP and 400d samples, 2) we compare the stacked profiles in the overlapping region of redshift and mass $0.5 < z < 0.9,\ 2.5 \times 10^{14} M_{\odot} < M < 5 \times 10^{14} M_{\odot}$. Fig. 2.7, 2.8 show the fitted emission measure, gas density and α for these 2 tests. The upper 4 panels in these plots are the same as in Fig. 2.5 – emission measure, emission measure in outskirts, gas density and α. The bottom 3 panels show deviations of emission measure, gas density and alpha from same quantities derived for the full stack (normalized by average values).

Both the full 400d and XVP samples, and their overlapping parts show a reasonable agreement between catalogs. The corresponding significances ζ are less than 2 for full catalogs and ~ 3 for their overlapping parts, which is slightly too high, and can be explained by random concentration of merging clusters in the overlapping part of two catalogs. We have not finished investigating this phenomenon, but we still believe that the difference in selection is not important for the 400d and XVP samples, and we combine the clusters in all following tests.

We note that we also address the question of similarity between 400d and XVP samples from a different perspective in Chapter 4.

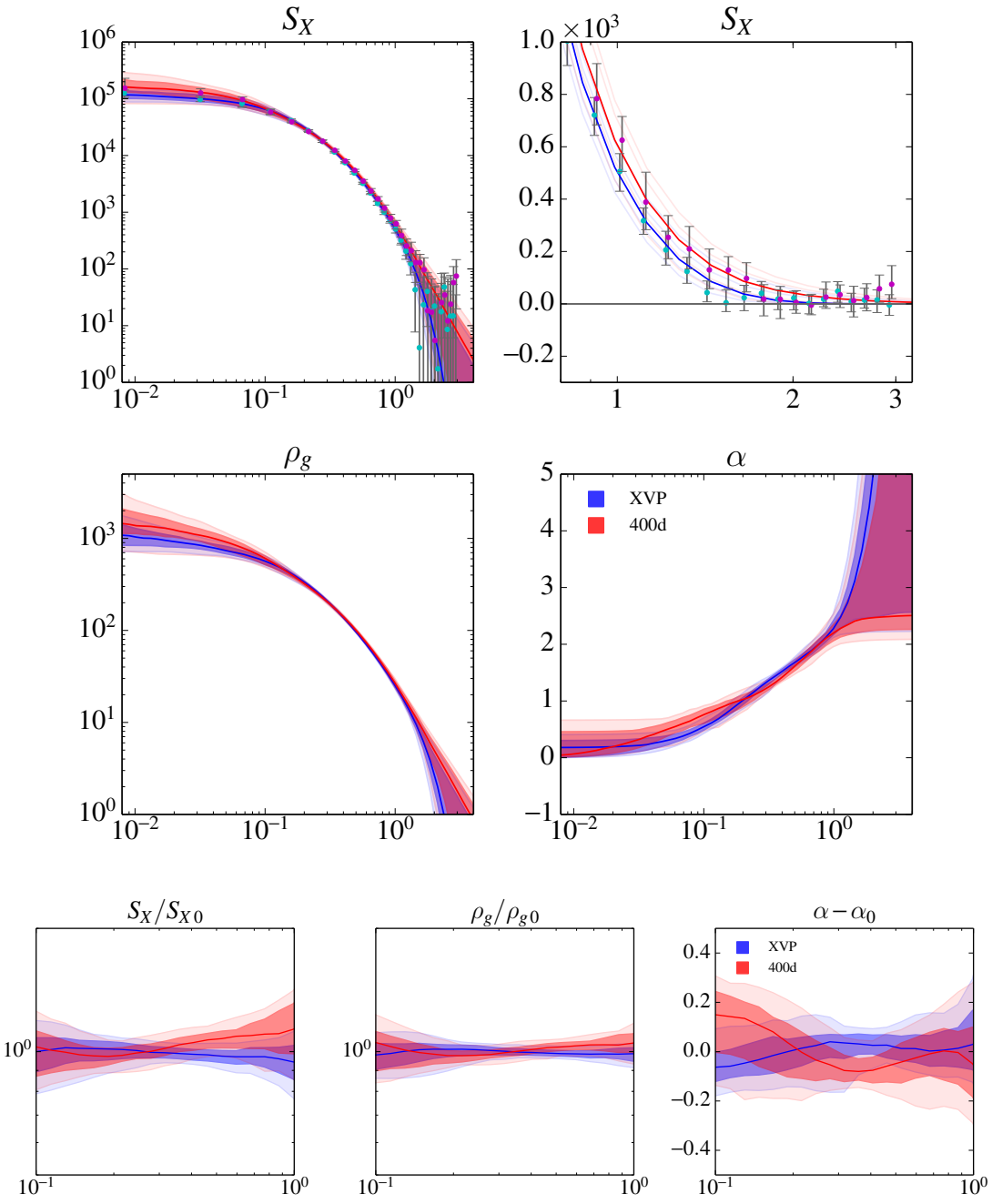

Figure 2.7: Comparison of XVP and 400d catalogs. The top 4 panels are the same as in Fig. 2.5, but now show both XVP (blue) and 400d (red) data points and fitted profiles. The bottom 3 panels show deviations of 400d and XVP average profiles from the profiles averaged over the entire catalog. These 3 panels, clearer than the top 4, show that confidence intervals for S_X, ρ_g and α overlap very well, and therefore we do not see a difference between average profiles in these catalogs.

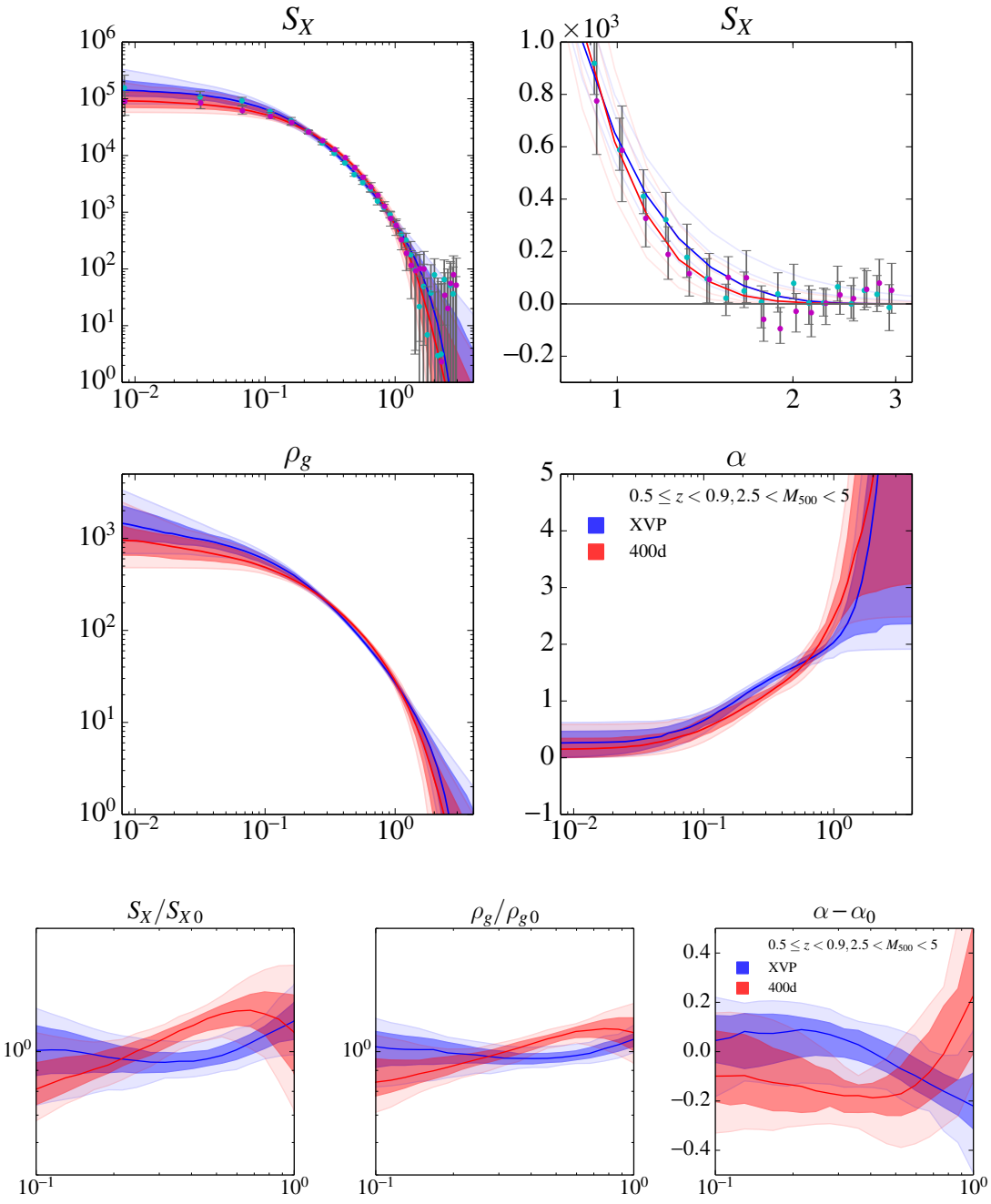

Figure 2.8: Comparison of overlapping parts of XVP and 400d catalogs. The top 4 panels are the same as in Fig. 2.5, but now show both XVP (blue) and 400d (red) data points and fitted profiles. The bottom 3 panels show deviations of the average profiles for the chosen subsamples from the profiles averaged over the entire catalog. These 3 panels, clearer than the top 4, show the difference between subsamples

2.5.3 Mass, redshift and $M/M_{\rm NL}$ bins

Cluster formation and evolution models predict that the shape of a cluster may depend on its mass (Kaiser, 1991), redshift (McDonald et al., 2013), and the mass normalized by the characteristic clustering mass $M_{\rm NL}$ (Vikhlinin et al., 2009; Kravtsov et al., 2005). We therefore separate the clusters into 3 bins by each of these observables. We choose the bin edges to have approximately equal number of clusters in each. The choice of bins is summarized in Table 2.1 and *the maximum* significance of difference between them is summarized in Table 2.2, the plots of ρ_g, α and normalized ρ_g and α are presented in Fig. 2.9-2.14.

The fitted models are presented in Fig. 2.9, 2.10, 2.11. We see some marginal differences between the bins, but they are too weak to say that we detected deviations from self-similarity correlated with the global cluster properties.

2.5.4 Central entropy

Our next test is partitioning the clusters by central entropy. Fig. 2.12 shows that although profiles differ in the central region (by construction), they are not significantly different beyond $0.3R_{500}$. There is an indication that NCC clusters have a higher normalization of gas density in the medium ranges of r ($0.3-1R_{500}$), as suggested by previous studies (Arnaud et al., 2002), but the evidence is not strong enough to claim a detection of this effect. A bigger sample is needed to make a decisive conclusion.

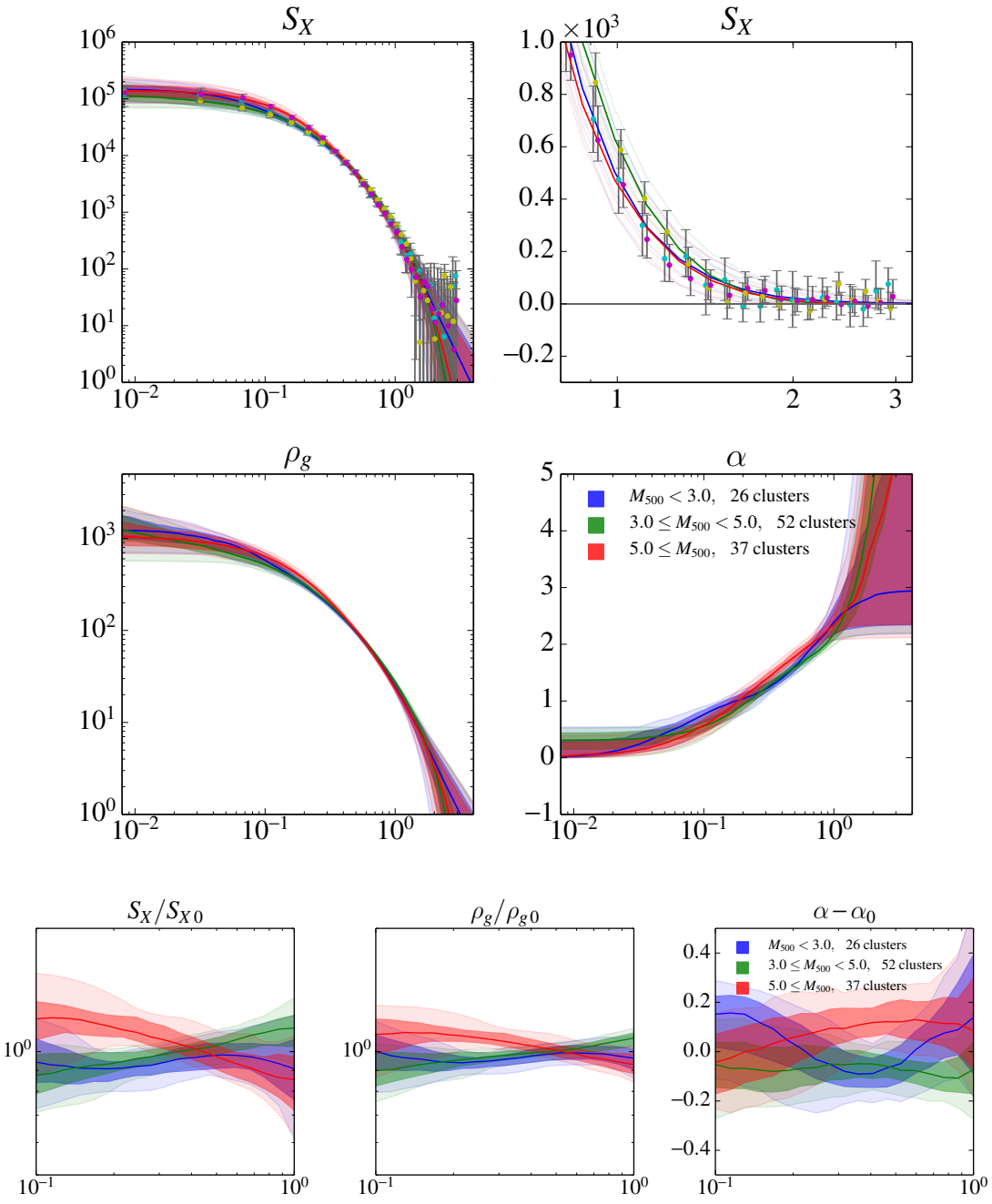

Figure 2.9: Mass bins. The definition of the bins is given in the plots and in Table 2.1. The top 4 panels are the same as in Fig. 2.5, but now show the 3 mass bins in red, green and blue. The bottom 3 panels show deviations of the average profiles in the chosen subsamples from the profiles averaged over the entire catalog. These 3 panels, clearer than the top 4, show the difference between subsamples

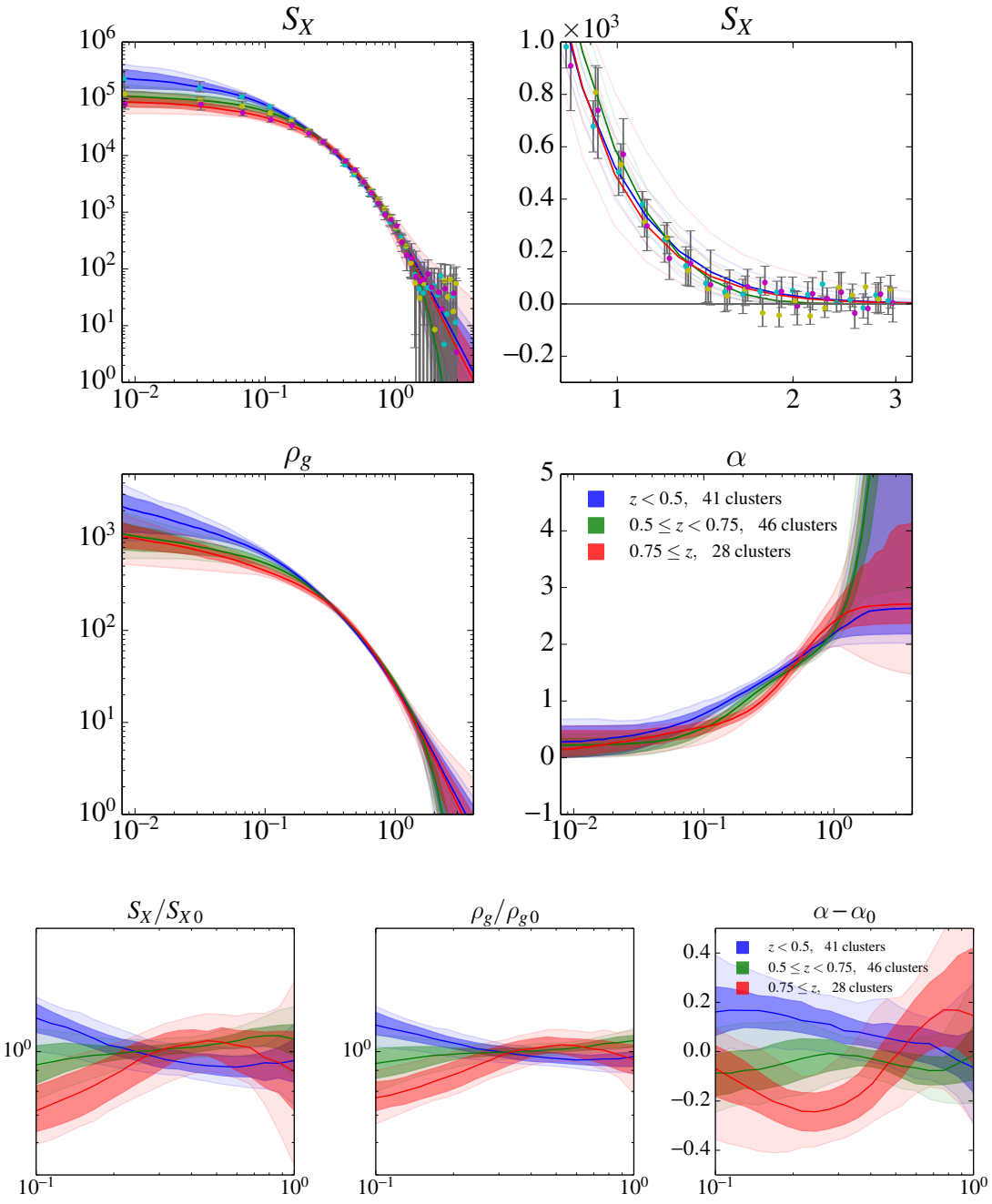

Figure 2.10: Redshift bins. The definition of the bins is given in the plots and in Table 2.1. The top 4 panels are the same as in Fig. 2.5, but now show the 3 redshift bins in red, green and blue. The bottom 3 panels show deviations of the average profiles in the chosen subsamples from the profiles averaged over the entire catalog. These 3 panels, clearer than the top 4, show the difference between subsamples

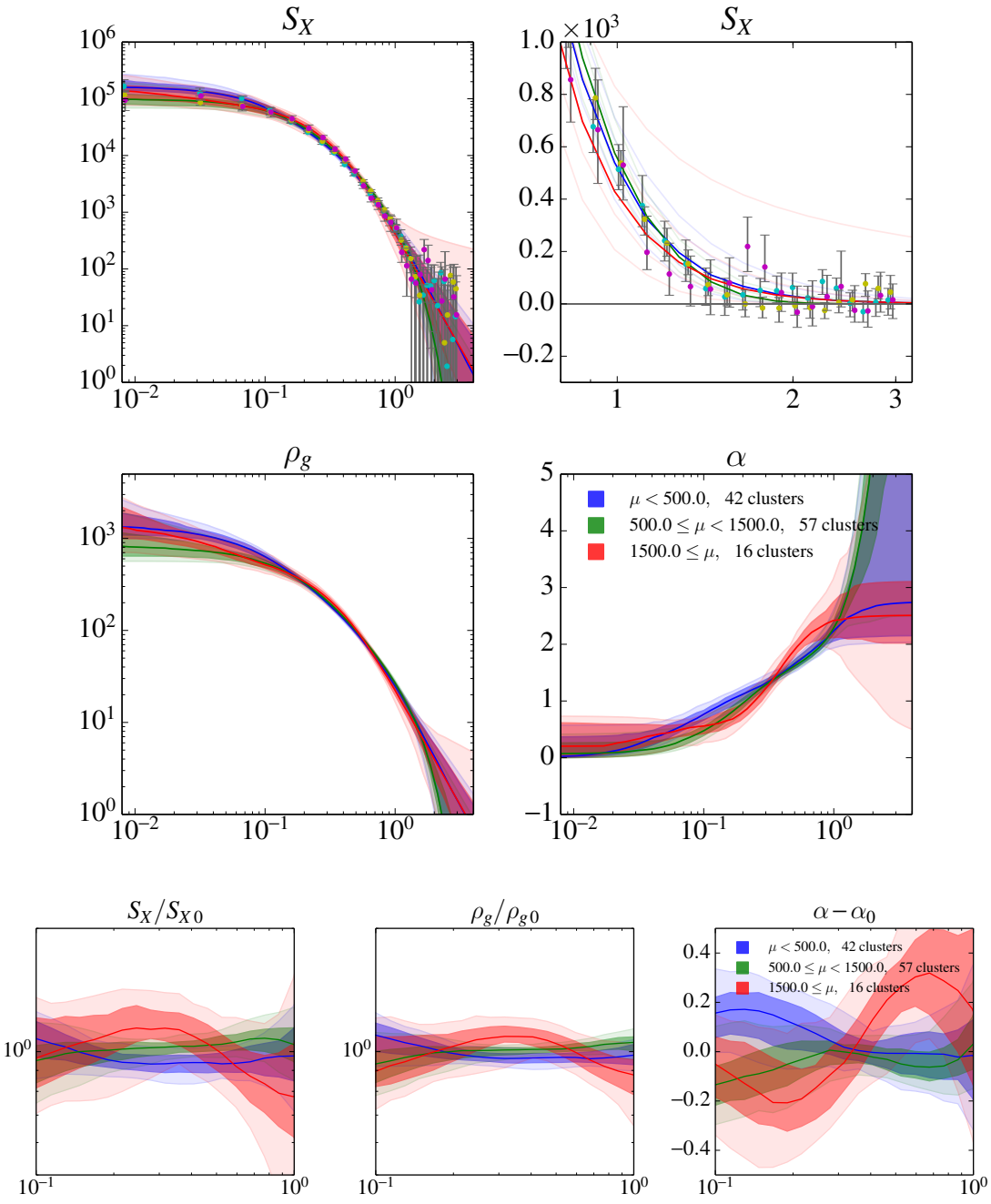

Figure 2.11: $\mu = M/M_{\rm NL}$ bins. The definition of the bins is given in the plots and in Table 2.1. The top 4 panels are the same as in Fig. 2.5, but now show the 3 $\mu = M/M_{\rm NL}$ bins in red, green and blue. The bottom 3 panels show deviations of the average profiles in the chosen subsamples from the profiles averaged over the entire catalog. These 3 panels, clearer than the top 4, show the difference between subsamples

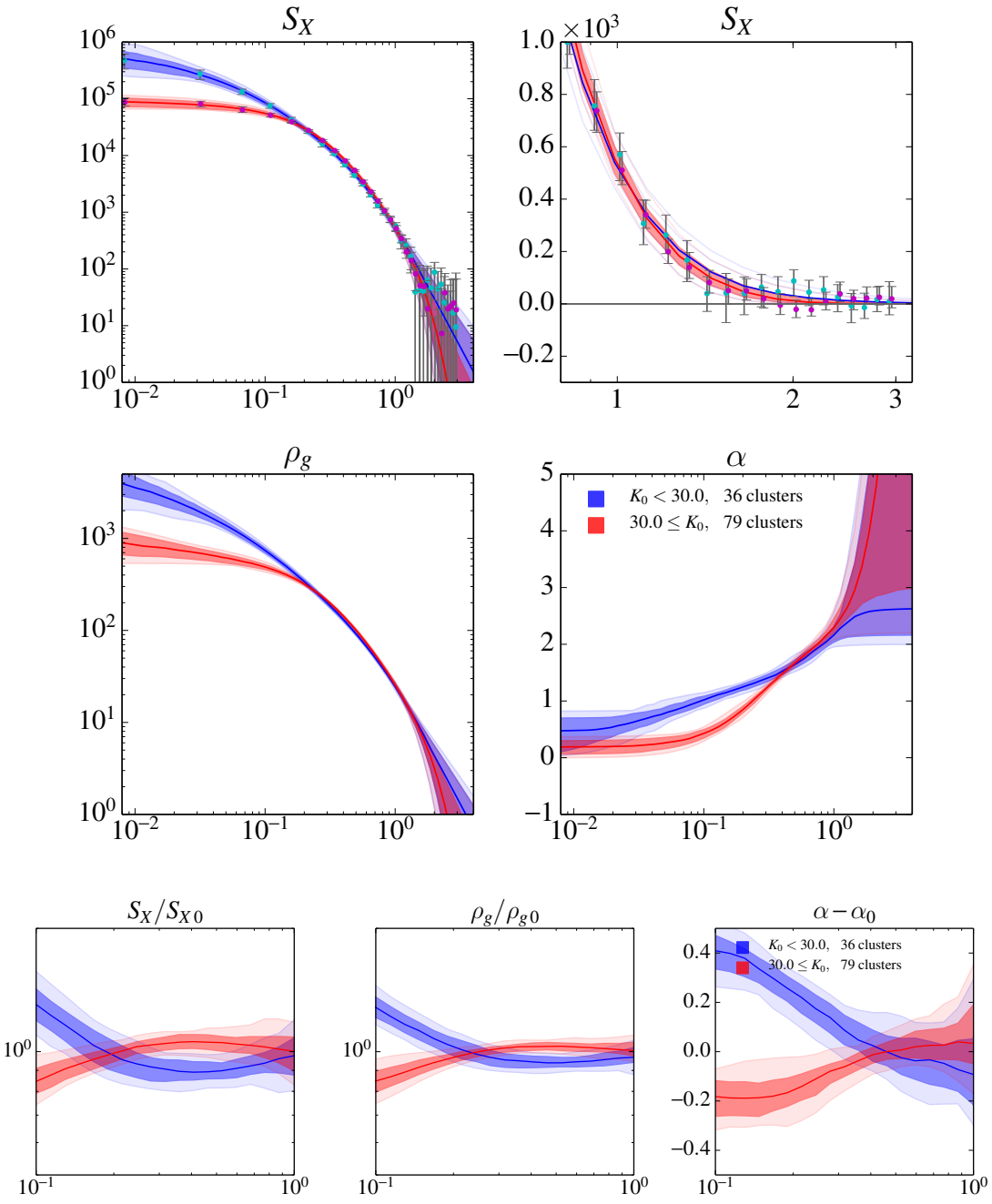

Figure 2.12: Central entropy bins. The definition of the bins is given in the plots and in Table 2.1. The top 4 panels are the same as in Fig. 2.5, but now show the 2 central entropy bins in red and blue. The bottom 3 panels show deviations of the average profiles in the chosen subsamples from the profiles averaged over the entire catalog. These 3 panels, clearer than the top 4, show the difference between subsamples

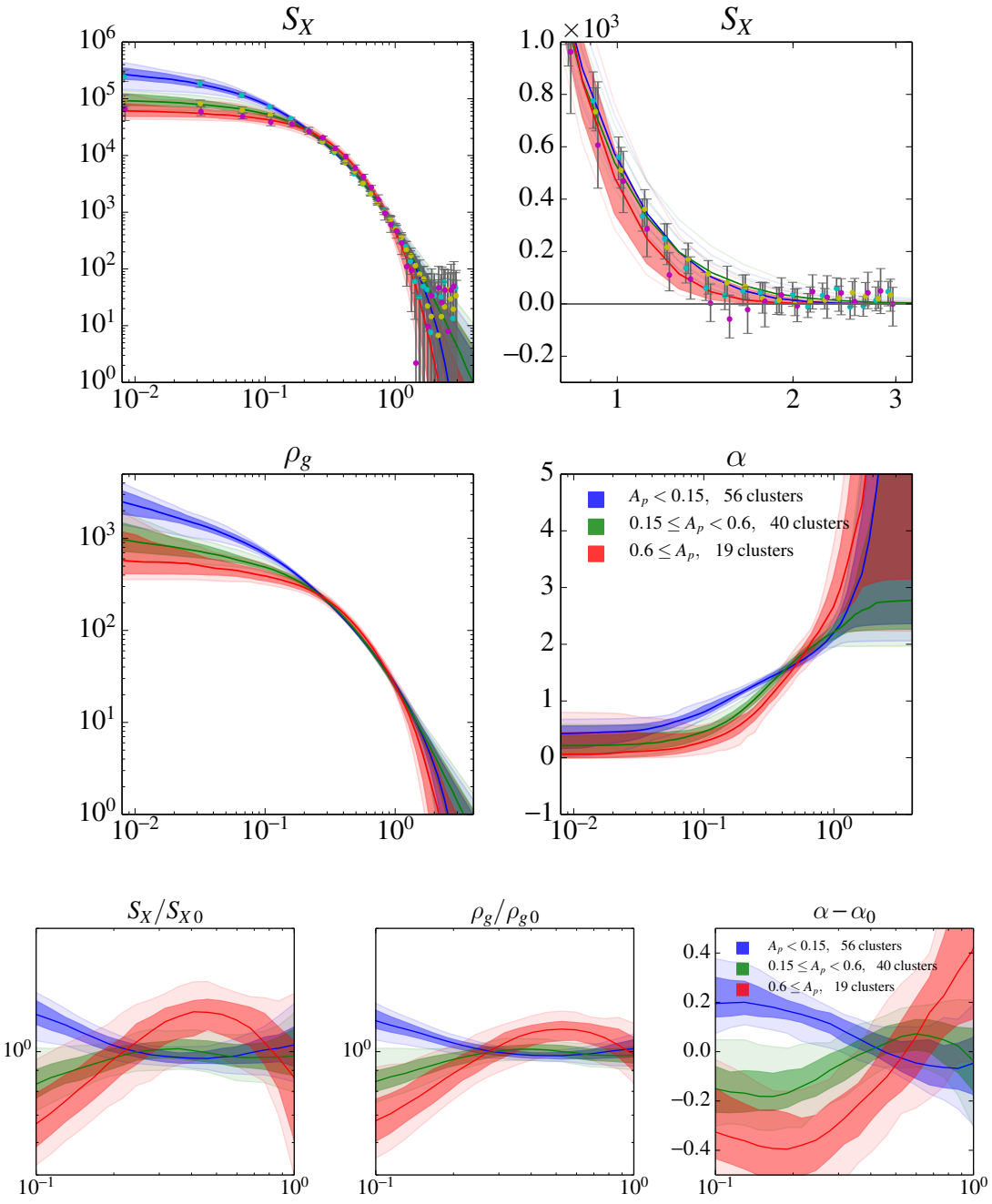

Figure 2.13: A_{phot} bins. The definition of the bins is given in the plots and in Table 2.1. The top 4 panels are the same as in Fig. 2.5, but now show the 3 A_{phot} bins in red, green and blue. The bottom 3 panels show deviations of the average profiles in the chosen subsamples from the profiles averaged over the entire catalog. These 3 panels, clearer than the top 4, show the difference between subsamples

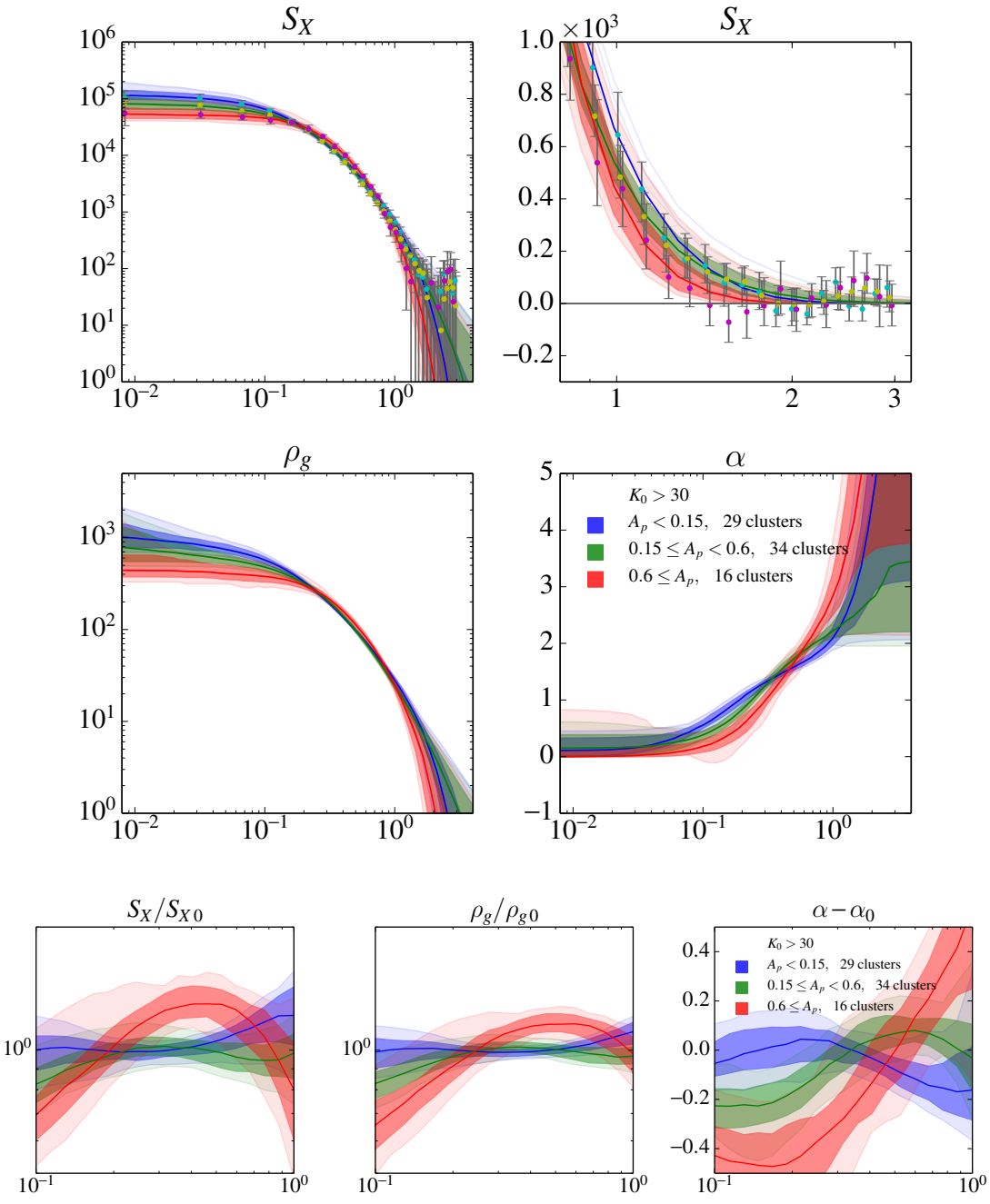

Figure 2.14: Non-cool cores A_{phot} bins. Same plot as the previous one (Fig. 2.13, but only including clusters without cooling flow as determined by central entropy ($K_0 > 30$ keV cm^2)

2.5.5 Dynamical state

It is believed that a cluster's dynamical state may be the most influential parameter causing scatter in the observables. (Krause et al., 2012; Ventimiglia et al., 2008; Kravtsov et al., 2006). We find that A_{phot} bins presented in Fig. 2.13 show the strongest deviations from self-similarity. Fig. 2.14 shows that dynamically disturbed clusters have bigger and flatter cores, higher gas density at 0.5 R_{500} and higher slope of gas density in the outskirts.

Since asymmetry anticorrelates with "cool coredness", it is possible that we only see the difference in A_{phot} bins because the lower bin contains most of the cool core clusters. To get rid of this effect, we repeated the comparison using NCC population only (as judged by central entropy). The new bins still show the same effects: steeper outer slope and higher gas normalization at ~ 0.5 R_{500}.

2.6 Conclusions

1. Galaxy cluster density profiles show a remarkable degree of self-similarity beyond cores (0.3 R_{500}) after proper cosmological scaling, as expected by purely gravitational models of structure formation (Kaiser, 1986, e.g.).
2. We detected X-ray emission beyond R_{500}.
3. Density profiles show marginal trends (deviations from self-similarity) with mass, redshift, M/M_{NL}.
4. The only strong trend that we observed is correlation with cluster dynamical state as measured by A_{phot} .
5. CC and NCC systems conform to the universal density profile beyond 0.3 R_{500}.

6. We found a unifying framework for constructing density profiles.

7. Both GNFW and Generalized beta model describe density profiles well.

8. Best fit models to the universal profile are

$$\begin{aligned} \rho_g &= 19.9(1+(r/0.628)^{1.174})^{-3.152}\rho_c \quad \text{(GNFW)} \\ \rho_g &= 401(r/0.179)^{-0.343}(1+(r/0.179)^2)^{-0.516}(1+(r/0.533)^3)^{-0.332}\rho_c \quad \text{(GB)} \end{aligned} \tag{2.32}$$

Chapter 3

Robust Quantification of Galaxy Cluster Morphology Using Asymmetry and Central Concentration

We present a novel quantitative scheme of cluster classification based on the morphological properties that are manifested in X-ray images. We use a conventional radial surface brightness concentration parameter (c_{SB}) as defined previously by others, and a new asymmetry parameter, which we define in this paper. Our asymmetry parameter, which we refer to as *photon asymmetry* (A_{phot}), was developed as a robust substructure statistic for cluster observations with only a few thousand counts. To demonstrate that photon asymmetry exhibits better stability than currently popular power ratios and centroid shifts, we artificially degrade the X-ray image quality by: (a) adding extra background counts,

(b) eliminating a fraction of the counts, (c) increasing the width of the smoothing kernel, and (d) simulating cluster observations at higher redshift. The asymmetry statistic presented here has a smaller statistical uncertainty than competing substructure parameters, allowing for low levels of substructure to be measured with confidence. A_{phot} is less sensitive to the total number of counts than competing substructure statistics, making it an ideal candidate for quantifying substructure in samples of distant clusters covering wide range of observational S/N. Additionally, we show that the asymmetry-concentration classification separates relaxed, cool core clusters from morphologically-disturbed mergers, in agreement with by-eye classifications. Our algorithms, freely available as *Python* scripts (`https://github.com/ndaniyar/aphot`) are completely automatic and can be used to rapidly classify galaxy cluster morphology for large numbers of clusters without human intervention.

3.1 Introduction

Clusters of galaxies are complex objects where many astrophysical processes are taking place. Cluster classification based on their X-ray morphology can help us understand the dominant physical processes in particular types of clusters, shed light on their formation histories, and give new insights into the evolution of both the large scale structure of the Universe (Allen et al., 2011) and the baryonic component of galaxy clusters (Böhringer & Werner, 2010).

Two distinctive features of galaxy clusters that are detectable in X-ray images are 1) cool cores and 2) departure from axial symmetry, presumed to arise from galaxy cluster mergers. Cool cores exhibit sharp central peaks in X-ray emission, while asymmetry manifests

as secondary peaks, filaments, and clumps in X-ray surface brightness. It is believed that these features emerge at different stages of cluster evolution, and are outcomes of completely different physical processes that affect the entire intracluster medium (ICM). One important reason to classify cluster morphology is that we can explore any correlations between morphology and residuals in various cluster scaling relations, resulting in more robust estimates of, for example, galaxy cluster mass (M_{500}).

The substructure clumps in the X-ray emission are often associated with active processes of dynamical relaxation after mergers. For such clusters (with a high amount of substructure) the characteristic processes are turbulence (Vazza et al., 2011; Hallman & Jeltema, 2011), shocks, and cold fronts in the ICM (Markevitch & Vikhlinin, 2007; Hallman et al., 2010; Blanton et al., 2011), giant and mini radio halos (Cassano et al., 2010) and relics (Ferrari et al., 2008). After the process of relaxation is over, cool cores start to develop (Fabian et al., 1994; Peterson & Fabian, 2006; Hudson et al., 2010; McDonald et al., 2013), and the evolution of the ICM is governed by the processes of gas cooling and heating, AGN feedback (McNamara & Nulsen, 2007) and thermal conduction (Voit, 2011). We are still far from a detailed understanding of these processes, but their correlation with morphology is established both from observations and simulations. For example, observations suggest that more dynamically disturbed systems have weaker cool cores (Sanderson et al., 2009).

In this work, we propose a new classification scheme, based on the arrangement of galaxy clusters in the 2-dimensional plane of *disturbance* and *cool core strength*. As explained above, this choice of fundamental morphological parameters is observationally well-motivated. To choose the parameters that best quantify cool core strength and disturbance, we first formulate some requirements:

1. These parameters need to be objective, quantitative and reproducible.
2. The parameters should be model independent.
3. They should allow substructure analysis for low signal-to-noise (S/N) observations.
4. These parameters should be relatively insensitive to exposure time, the level of the X-ray background or a cluster's angular size on the sky. A composite test that checks all these sensitivities together is simulating observations of a cluster at higher redshift.
5. The substructure parameters should agree with the human expert judgement.

The radial surface brightness profile of X-ray emission can be used to quantify the extent to which a cool core is present, although assigning clusters to categories (cool core vs. non cool core) is still a topic of discussion (Hudson et al., 2010; McDonald et al., 2013). We adopt here the concentration prescription of Santos et al. (2008), who showed that their implementation can discriminate between "strong", "medium" and "no" cool cores. Importantly, in the context of the requirements listed earlier, the Santos et al. (2008) concentration parametrization is robust even for low S/N observations and is roughly model independent.

The quantification of "disturbance" is significantly harder. There is no simple physical (or mathematical) quantity that can measure "disturbance" or as it is usually called, the amount of substructure. Pinkney et al. (1996) list 30 different substructure tests and conclude that no single one is good in all cases. Two substructure statistics have, nevertheless, became popular recently: centroid shifts (Mohr et al., 1993) and power ratios (Buote & Tsai, 1995, 1996). Their popularity can be explained by their model-independence and ease of computation. They also satisfy, reasonably well, the requirements formulated above. For a more detailed review of various substructure statistics see Buote (2002); Böhringer

et al. (2010); Rasia (2013); Weißmann et al. (2013b). We present below a new substructure statistic that is superior based on the above requirements.

We stress that any substructure statistic should be suitable for high redshift clusters, with observations of poor quality. This is an area where the other substructure tests do not perform very well. Most morphological studies have been carried out for nearby clusters with high S/N X-ray images (10^5 counts per cluster being the typical value for these studies). However, large surveys or serendipitous discoveries of high-redshift clusters will yield images with, typically, only several hundred counts (e.g., McDonald et al., 2013). Thus, a reliable next-generation substructure statistic must perform equally well on low-S/N, high-redshift observations.

Here, we present a new substructure statistic, *photon asymmetry* (A_{phot}), which quantifies how much the X-ray emission deviates from the idealized axisymmetric case. This statistic is somewhat similar to existing efforts to use the residuals after subtracting a beta-model fit (e.g. Neumann & Bohringer, 1997; Böhringer et al., 2000; Andrade-Santos et al., 2012) or double beta-model (Mohr et al., 1999). However, A_{phot} is model independent and specially designed to work well for observations with low photon counts.

In §2, we present the X-ray sample that has been used to develop our approach. §3 defines the various morphology measures that are compared in this work, while §4 explores performance using simulated data sets. Our results and conclusions are presented in §5 and §6, respectively. We defer an analysis of how these morphological parameters correlate with the scaling relations residuals to a future publication.

3.2 Sample and data reduction

3.2.1 Sample

To test our classification method and compare the properties of photon asymmetry to the properties of previously known substructure statistics, we used the high-z subsample of the 400 square degree galaxy cluster survey (abbreviated as 400d), which is a quasi mass-limited sample of galaxy clusters at $z > 0.35$ serendipitously detected in ROSAT PSPC data (Burenin et al., 2007). The high-redshift subsample of 400d was published in Vikhlinin et al. (2009a) and consists of 36 clusters with $z > 0.35$ which exceed a certain luminosity threshold which corresponds to $\approx 10^{14} M_{\odot}$ in mass (see Vikhlinin et al. (2009a) for details).

All clusters in the sample have been observed with the *Chandra X-ray Observatory* and used to constrain cosmological parameters in Vikhlinin et al. (2009a,b).

The reasons for choosing this cluster sample are:

1. A redshift range that covers $0.3 \leq z \leq 0.9$, and is similar to the redshift range of both SZ surveys and next-generation X-ray surveys (e.g., eRosita), allowing extension to larger samples in the future.

2. High-resolution Chandra imaging which is very suitable for substructure detection. As we show in the paper, telescope resolution is very important to detect and quantify substructure.

3. A range of photon counts. Since our goal is to develop a substructure statistic that is maximally applicable to high-z clusters with low S/N observations, the high-z part of the 400d catalog is perfect for testing our substructure statistic.

4. The basic selection criterion is X-ray luminosity, which adequately samples the range

of cluster morphologies and core properties. Thus the sample should be representative with respect to cluster morphological types.

3.2.2 Data reduction

We perform all industry-standard X-ray data reduction steps. We start with flare cleaned event2 files that are identical to those used by Vikhlinin et al. (2009a). Following many other cluster studies (e.g. Santos et al., 2008), we apply a 0.5 - 5.0 keV band filter which optimizes the ratio of the cluster to background flux. We chose to use a higher upper cut-off than what was used in many other studies (2 keV), because for massive clusters there is significant emission above 2 keV.

We detect point sources with an algorithm similar to *wavdetect* from the CIAO package (Fruscione et al., 2006) and replace the regions of point sources with a Poisson distribution with a mean value equal to the local background density of counts. In most cases this means that we add no counts in the region of the removed point source because the typical local background level is $\sim 10^{-2}$ counts per pixel.

We estimate the global background level from regions on the chip free of point sources, away from chip gaps, and sufficiently far away from cluster center (2-4 R_{500} annulus)

We compute all morphological parameters directly from the raw event2 band-filtered files without additional binning or smoothing. All substructure statistics that we consider in this paper can be formulated in terms of sums over counts instead of integrals over surface brightness distributions as they are usually presented. We believe that this is the best way to perform statistical tests because any post-processing may distort and bias the statistic's distributions.

We use exposure maps that include corrections for CCD gaps, spatial variations of the effective area, ACIS contamination, bad pixels and detector quantum efficiency.

We produce smoothed images of the clusters using an algorithm similar to *asmooth* (Ebeling et al., 2006), which chooses the appropriate smoothing scale adaptively for each count based on the local density of counts. These smoothed images are used for two (and only two) purposes:

1. Visualization for by-eye classification and by-eye comparison of the cluster's relative ranking produced by various substructure statistics,
2. Generation of simulated cluster observations. See Section 3.5.2 for more details.

All the steps in the data reduction pipeline are automatic, but the results of each step were visually inspected. For the clusters that had several observations, we merged all observations that had the entire R_{500} aperture on the CCD.

3.3 Classical morphological parameters/ substructure statistics

3.3.1 Power ratios

Power ratios were introduced in Buote & Tsai (1995, 1996) and have been widely used ever since (e.g. Jeltema et al., 2005; Ventimiglia et al., 2008; Cassano et al., 2010). They are able to distinguish a large range of morphologies, physically motivated and easy to compute (Jeltema et al., 2005). The method consists of a multipole expansion of the surface brightness and computes the powers in different orders of the expansion. The corresponding

formulas are usually quoted as integrals over surface brightness, but since we prefer to work with individual counts and not smooth the surface brightness in any way, we replaced all the integrals with appropriately weighted sums over counts.

The powers are given by:

$$P_0 = [a_0 \ln(R_{ap})]^2 \tag{3.1}$$

$$P_m = \frac{1}{2m^2 R_{ap}^{2m}} \left(a_m^2 + b_m^2\right), \tag{3.2}$$

where R_{ap} is the aperture radius. The moments a_m and b_m are calculated using

$$a_m(R) = \sum_{r_i \leq R_{ap}} w_i r_i^m \cos(m\phi_i) \tag{3.3}$$

and

$$b_m(R) = \sum_{r_i \leq R_{ap}} w_i r_i^m \sin(m\phi_i), \tag{3.4}$$

where r_i, ϕ_i are the coordinates of the detected photon in polar coordinates and w_i is its "weight" which is inversely proportional to the effective exposure at the given CCD location. The center of that polar coordinate system is chosen to set P_1 to zero.

In order to render the morphological information insensitive to overall X-ray flux, each of the angular moments $P_m, m = 1, 2, 3...$ is normalized by the value of P_0, forming the power ratios, P_m/P_0. The power ratios P_2/P_0, P_3/P_0, P_4/P_0 have been used to characterize cluster substructure (Jeltema et al., 2005). P_3/P_0 has been found to be the best characterization of "disturbance".

Aperture choice is very important for power ratios as they are most sensitive to the substructure at the maximum radius. Values of 1 Mpc, 0.5 Mpc, R_{500} have been used as aperture radii. We use R_{500} as it allows more consistent comparison of clusters of different mass than a fixed physical scale as R_{500} is a natural scale for clusters of all masses and

redshifts. The other substructure statistics are also based on an R_{500} aperture, therefore, our comparison of various substructure statistics is consistent.

As many authors have noted, the power ratios calculated by the formulas above give values for P_m biased high due to photon noise. This can be easily seen in the case of a perfectly symmetrical cluster - the random distribution of the angles ϕ_i, and nonnegativity of P_m, lead to a distribution of P_m with nonzero mean. Different authors used different methods to account for these biases. We based our method of bias correction on the work of Böhringer et al. (2010), where the bias was computed by randomizing the polar angles for all collected photons, but keeping their radial distance fixed. The mean of the power P_m of the mock observations obtained this way is interpreted as the typical photon noise contribution to the measurements of P_m and subtracted from the P_m of the real observation. We did not perform Monte-Carlo simulations for randomizing polar angles, because the mean of P_m with randomized angles (uniformly distributed ϕ_i) can be easily calculated analytically:

$$\begin{aligned} n_m = \langle a_m^2 \rangle &= \left\langle \left[\sum w_i r_i^m \cos(m\phi_i) \right]^2 \right\rangle \\ &= \sum w_i^2 r_i^{2m} \langle \cos^2(m\phi_i) \rangle = \frac{1}{2} \sum w_i^2 r_i^{2m} \end{aligned} \tag{3.5}$$

We need to subtract this value from both a_m^2 and b_m^2 which results in the following formula for P_m.

$$P_m = \frac{1}{2m^2 R_{ap}^{2m}} \left(a_m^2 + b_m^2 - 2n_m \right), \tag{3.6}$$

After bias correction the background counts do not contribute to the powers $m \neq 0$, but still contribute to $a_0 = \sum w_i$. To make P_0 and, consequently the ratios background independent, we need to also subtract the background contribution from a_0:

$$a_0 = \sum w_i - w_{\mathrm{bkgr}}(R_{ap}), \tag{3.7}$$

where $w_{bkgr}(R_{ap})$ is the expected total weight of all background photons within the aperture R_{ap}.

3.3.2 Centroid shifts

Centroid shift is another popular measure of "disturbance" of clusters. It is defined by the variance of "centroids" obtained by minimization of P_1 within 10 apertures($r \leq n \times 0.1\, R_{500}$, with n = 1,2..10). The value of centroid shifts is expressed in units of R_{500} which makes it a dimensionless quantity. Centroid shifts are defined slightly differently by different authors (See Mohr et al., 1995; Poole et al., 2006; O'Hara et al., 2006; Böhringer et al., 2010). Here we used

$$w = \left[\frac{1}{N-1} \sum_i \left(x_i - \langle x \rangle \right)^2 \right]^{1/2} \times \frac{1}{R_{500}}, \tag{3.8}$$

where x_i is the position of the centroid of a given aperture.

3.3.3 Concentration

Concentration parameter is defined as the ratio of the peak over the ambient surface brightness. Concentration has been widely applied to X-ray images (Kay et al., 2008; Santos et al., 2008, 2010; Cassano et al., 2010; Hallman & Jeltema, 2011; Semler et al., 2012) and proved useful in distinguishing cool-core (CC) from non-cool-core (NCC) clusters.

We adopted the definition of concentration provided by Santos et al. (2008):

$$c_{SB} = \frac{\text{Flux}(r < 40\text{kpc},\ 0.5\text{keV} < E < 5\text{keV})}{\text{Flux}(r < 400\text{kpc},\ 0.5\text{keV} < E < 5\text{keV})} \tag{3.9}$$

The radii 40 and 400 kpc were chosen to maximize the separation between CC and NCC clusters. We computed concentration around the brightness peak as defined in Section 3.4.4,

the same center that we used for photon asymmetry. Complete details on the stability of the concentration parameter can be found in Santos et al. (2008).

3.4 Photon asymmetry

In this section we will describe our proposed morphological classifier, namely photon asymmetry.

3.4.1 Optical asymmetry and the motivation for photon asymmetry

In optical astronomy the asymmetry parameter is a part of the "CAS" galaxy classification scheme which stands for concentration (C), asymmetry (A) and clumpiness (S) (Conselice, 2003). Asymmetry quantifies the degree to which the light of an object (galaxy) is rotationally symmetric. It is measured by subtracting the galaxy image I_{180} rotated by 180° from the original image I_0 (Conselice, 2003):

$$A = \frac{|I_0 - I_{180}|}{I_0} \tag{3.10}$$

This definition tests *central* (or *mirror*) asymmetry, i.e. whether the image is invariant under a "point reflection" transformation (which is equivalent to rotation by 180° around the central point). Although this definition of asymmetry has been applied to X-ray images of clusters before (e.g. Rasia, 2013), it is only reliable for observations where the number of counts in each (binned) pixel is $\gg 1$. This condition is not satisfied for most cluster observations.

One can come up with a similar definition of *circular* or *axial* asymmetry which would test whether the image is invariant under rotation by arbitrary angle around the central point. That would involve finding the average intensity of the image in concentric annuli, and comparing local intensity with the average intensity in the annulus.

$$A = \int_0^R r dr \int d\phi \left(I(r,\phi) - \overline{I(r)} \right)^2 \tag{3.11}$$

This could also be a good measure of substructure and indeed people have tried to apply similar ideas for substructure statistics (e.g. Andrade-Santos et al., 2012).

The above definitions of asymmetry, both (3.10) and (3.11), are hard to implement for distant clusters whose observations have fewer counts. We could generate smoothed images of clusters and apply the above definitions to these images, but that can generate biases. The large radial variations in surface brightness and the presence of substructure deny the ability to choose a single, global optimal smoothing scale. We cannot use an adaptive scale either, because asymmetry is then strongly dependent on the details of the adaptive smoothing algorithm. Also, by producing smoothed images (with either a fixed or adaptive scale), we effectively introduce some model-dependent prior on cluster structure. We would prefer, however, to only use objective information: the positions (and possibly energies) of the detected photons.

Fortunately, there is a way to adapt the definition of asymmetry so that it can be computed efficiently in the limit of low photon counts, which we present in this paper. This adaptation is possible for both *central* and *axial* asymmetry. Central asymmetry might seem preferable, because it would have a zero value for a relaxed, but elliptical cluster. However, in our sample with few counts and ill-defined ellipticities, the values of axial and central asymmetries correlate strongly. Additionally, *axial* asymmetry is conceptually simpler for

our statistical framework, so we concentrate on it for this paper.

Our strategy for adapting Eq. (3.11) to the case of few counts with known coordinates is the following. We split the image into a few annuli, and check whether the surface brightness is uniform in each of these annuli. In the limit of few counts, this is the same as checking whether these counts are uniformly distributed in the annulus. This amounts to checking that their polar angles are uniformly distributed in the $0 \leq \phi < 2\pi$ range.

3.4.2 Photon asymmetry within an annulus

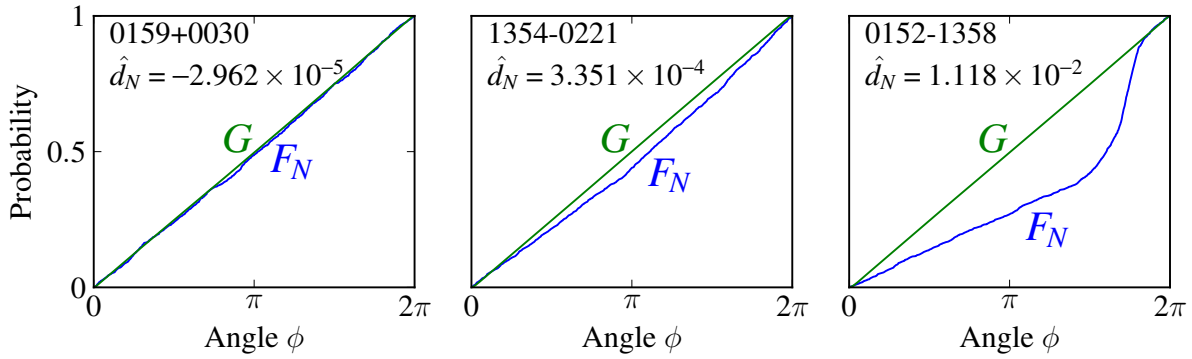

Figure 3.1: The method described in this paper compares the observed cumulative probability distributions of the angular positions of photons (F_N), to that of a uniform distribution (G). This Figure shows empirical F_N and uniform G distribution functions in the outermost annulus for 3 progressively more disturbed clusters (0159+0030, 1354-0221, and 0152-1358). The more disturbed clusters manifest greater differences between F_N and G and correspondingly higher values of distance metric d_N.

To assess the degree of nonuniformity of the angular distribution of the counts, we use Watson's test (Watson, 1961). Watson's test compares 2 cumulative distribution functions. Other members of this family of non-parametric tests for the equality of distribution functions include the well-known Kolmogorov-Smirnov test as well as less well-known Cramer-von Mises and Kuiper's tests. For the reasons explained in the Appendix, Watson's test is the

only one that works in our specific situation. Unfortunately, Watson's test is only able to test the null hypothesis, i.e. compute the probability that the given sample is drawn from the assumed distribution. Our case is slightly different - we know that our sample (of counts as function of polar angles) is not drawn from the uniform distribution, so, in principle, goodness of fit tests are not applicable to our case. However, as we show in the Appendix, we can interpret the value of Watson's test as the estimate of the distance between the true underlying distribution function and the assumed distribution function.

Let us consider the photons that arrive in an annulus $R_{in} < r < R_{out}$ relative to the cluster center. The specific definition of these annuli will be discussed in Section 3.4.3. Let Φ be a polar angle (random variable) of a cluster photon in the chosen coordinate system, centered on the cluster, and $\phi_1, \phi_2, \cdots, \phi_N$ are the polar angles of the observed photons in the annulus (N = total number of observed photons in the annulus). Then we will define:

$$F(\phi) = \text{Prob}(\Phi \leq \phi) \tag{3.12}$$

as the true angular (cumulative) distribution function and

$$F_N(\phi) = \frac{1}{N} \sum \begin{cases} 1, & \text{if } \phi_i \leq \phi \\ 0, & \text{otherwise} \end{cases} \tag{3.13}$$

as the measured (empirical) distribution function. Being distribution functions on a circle, F and F_N also depend on the arbitrary starting point ϕ_0 which we write as

$$\begin{aligned} F &= F(\phi; \phi_0) \\ F_N &= F_N(\phi; \phi_0). \end{aligned} \tag{3.14}$$

We can now introduce Watson's statistic $U_N^2[F_N, F]$ as

$$\begin{aligned} U_N^2[F_N, F; \phi_0] &= N \int \Big(F_N(\phi; \phi_0) - F(\phi; \phi_0)\Big)^2 dF \\ U_N^2[F_N, F] &= \min_{\phi_0} U_N^2[F_N, F; \phi_0], \end{aligned} \tag{3.15}$$

i.e. U_N is the minimum value of integrated squared difference between F_N and F over possible starting points ϕ_0.

The greater the value of U_N^2 the less likely that F_N is produced by drawing from F. In our case F is unknown, but we can test how likely it is that F_N is drawn from another distribution G which represents an idealized axisymmetric source. (G would be uniform in the absence of instrumental imperfections)

$$U_N^2[F_N, G] = N \min_{\text{origin on the circle}} \int (F_N - G)^2 dG. \tag{3.16}$$

Interestingly, it is possible to interpret $U_N^2[F_N, G]/N$ as the distance between F and G:

$$U_N^2[F_N, G]/N = \text{distance}(F, G) + \frac{1}{12N} + \text{Noise}, \tag{3.17}$$

where the bizarre term $1/12N$ comes from the properties of the statistic distribution under the null hypothesis. The detailed derivation of (3.17) is presented in the Appendix. Here we would like to note that the mean value of Noise is smaller than $1/12N$ for the relevant values of N, therefore

$$\hat{d}_N = U_N^2/N - 1/12N \tag{3.18}$$

is an estimator of distance(F, G), the distance between the observed and uniform distributions of photons in the annulus. The variance of this estimator scales as $1/N$, so that we can get better estimate of the distance as N increases.

The method is illustrated in Fig. 3.1 where we show F_N, G, and the value of $\hat{d}_N$ in the outermost annulus for 3 progressively more disturbed clusters. The more disturbed clusters manifest greater differences between F_N and G and, consequently, higher values of $\hat{d}_N$.

As we are interested in the distance between the observed and a uniform distribution of *cluster-only* photons (as opposed to *cluster and background* photons), we additionally need to multiply that distance by the squared ratio of total counts N to cluster counts C in that annulus. As the number of cluster counts C is not directly observable, we estimate it by subtracting the expected number of background counts in the annulus from the total counts N. The resulting background-corrected expression

$$\hat{d}_{N,C} = \frac{N}{C^2}\left(U_N^2 - \frac{1}{12}\right), \tag{3.19}$$

is an estimate of the distance between the true photon distribution and the uniform distribution. (see Appendix for details)

3.4.3 How to choose the optimal annuli

The first step in choosing optimal annuli is to select the maximum aperture radius. R_{500} is a good choice, because cluster X-ray emission is typically indistinguishable from the background beyond this radius. Also, we exclude the region $r < 0.05R_{500}$ from the analysis because pixelation artefacts at small radii distort Watson's statistic.

Second, we need to choose the number of annuli inside this $0.05R_{500} < r < R_{500}$ region. One can use any number of annuli for the computation of asymmetry. The tradeoff is between the asymmetry S/N in each individual annulus and radial resolution. We found that it is desirable to have at least a few hundred counts in each annulus, so we used 4 annuli for our sample of clusters. This optimization may be different for a cluster sample

with different number of counts.

Finally, we need to choose the radii of these annuli. The relative uncertainty of asymmetry is estimated to be $\sqrt{N}/C$, where N is the total number of counts, and C is the number of cluster counts. The radial binning should be chosen carefully, because low numbers of N or C in any bin inflate the uncertainty. This is a nontrivial task as the radial brightness profile is very different between clusters. We choose the radial binning to achieve a uniform relative uncertainty in asymmetry, for each annulus, across the clusters in our training set. The following choice of boundaries in units of R_{500}: 0.05, 0.12, 0.2, 0.30, 1, leads to the most uniform uncertainty across bins. We caution that applying this technique to X-ray survey instruments or data sets that exhibit a broader PSF than Chandra's 0.5 arcsec FWHM will require a careful reevaluation of radial binning, sence we desire annuli widths $dr \gg$ FWHM.

The last step in the computation of photon asymmetry is to combine the values of asymmetry from the 4 annuli. We use a weighted sum of distances from each annulus $\hat{d}_{N_k,C_k}$ (see Eq. (3.19), k numbers the annuli, N_k and C_k are the total and cluster counts in k-th annulus) with a weight equal to the estimated number of cluster counts C_k in that annulus.

$$A_{phot} = 100 \sum_{k=1}^{4} C_k \hat{d}_{N_k,C_k} / \sum_{k=1}^{4} C_k \tag{3.20}$$

We introduced a multiple of 100 into the definition of $\mathrm{A_{phot}}$ to bring all the asymmetries to a convenient range $0 < \mathrm{A_{phot}} \lesssim 3$. The resulting quantity is independent of exposure and background level.

3.4.4 Cluster centroid determination

The standard prescription for optical asymmetry is to choose the center that minimizes asymmetry. However this method is prone to producing values of asymmetry that are biased

low. This effect is especially noticeable in our resamplings with very low number of counts.

We based our choice of centroiding on three considerations: 1) we favor a centroid choice that is independent of the asymmetry computation, 2) if the cluster possesses a strong core, we use that feature to define the cluster center, and 3) by assigning the cluster center to a high S/N region of the image, we can compute asymmetry in annuli at high S/N.

Based on these requirements, we chose the center to be the brightest pixel after convolution with a Gaussian kernel with $\sigma = 40$ kpc. At $z = 1$, a single Chandra pixel corresponds to about 4 kpc, and the Chandra PSF FWHM is of order 2 pixels, so the smoothing scale is much coarser than Chandra angular resolution.

The centroid defined as a convolution with a Gaussian kernel is not very sensitive to the size of this Gaussian kernel. We chose the kernel size to be 40kpc to be consistent with our definition of concentration. We use this centroid for both asymmetry and concentration. We stress that the Gaussian-convolved image is used only for centroiding, not for computation of any substructure statistics.

3.4.5 Additional remarks

The A_{phot} parameter can be applied beyond X-ray observations. In fact, a quantity defined exactly the same way can be used for optical observations of clusters if we replace the coordinates of each photon with the coordinates of a member galaxy. Additionally, each galaxy can be given a weight that depends on its optical luminosity (e.g. w $\sim$ log luminosity). These weights can replace the ones in eq.(13), so that F remains a proper distribution function. In this case the equations (13)-(20) remain valid yielding a parameter that describes the asymmetry of the galaxy distribution within a cluster derived from optical

observations.

3.5 Simulated observations and determination of uncertainties

3.5.1 Simulated observations

We now address the questions of 1) sensitivities of substructure statistics to observation parameters, and 2) uncertainties of these substructure statistics, by calculating them for simulated observations with the desired parameters (such as exposure or background level). The idea of using simulated observations in similar ways goes back to the works of Buote & Tsai (1996); Jeltema et al. (2005); Hart (2008); Böhringer et al. (2010). Generating these simulated observations is straightforward if we have the map of the true cluster surface brightness (or, more precise, cluster brightness multiplied by CCD exposure map) - we would draw each pixel value from the Poisson distribution with the mean equal to that brightness. As we don't know that true underlying brightness distribution, we use instead our best approximation to it, which is the result of an adaptive smoothing algorithm.

To simulate changing the exposure, before drawing from Poisson distribution, we need to multiply the surface brightness map by a constant; to change the level of background, we need to add a constant to the surface brightness map; to change the telescope PSF, we need to convolve the existing brightness map with the new PSF (the real Chandra PSF is negligibly small).

To simulate how the clusters would look if they were moved to a greater redshift, we

need to calculate the expected X-ray flux from that cluster, rescale the number of observed counts accordingly, change the image spatial scale (which is a small correction as angular diameter distance doesn't change much from z = 0.3 to 1), and then increase the amount of the background to its old value. The only tricky part in this process is the calculation of the new cluster flux which should include the change in the luminosity distance, and the K-correction (Hogg et al., 2002) that compensates for the shift in the cluster emission in the observed frame.

$$\frac{\mathrm{Flux_{new}}}{\mathrm{Flux_{old}}} = \frac{D^2_{L,\mathrm{old}}}{D^2_{L,\mathrm{new}}} \frac{K(z_{\mathrm{new}})}{K(z_{\mathrm{old}})} \tag{3.21}$$

Since we don't need to simulate this very precisely – we only want to get an idea of how it affects the substructure measures – we use a simple approximation to Santos et al. (2008) results for 0.5-5 keV energy band:

$$K(z) = \frac{1}{1+2z} \tag{3.22}$$

3.5.2 Uncertainties

To estimate the uncertainties of the various substructure statistics, we used the above-described algorithm to generate 100 mock observations with exactly the same exposure and background level as in the original observations, but varied noise realization. Then we computed the substructure statistics for these samples, and found the median, the 16th lowest and the 16th highest observed value in the sample. We treat the median as the characteristic central value of statistic for this set of mock observations, and the interval between 16th lowest and the 16th highest observed value as the 1σ, or 68% confidence interval. Using order statistics for the central value and the confidence interval is the most sensible choice for us, because the distributions for any substructure statistic values are asymmetric and

extremely heavy tailed. The statistic value obtained from the real observation didn't always fall within this confidence interval for two reasons. First, as this is only a 68% confidence interval, we expect approximately 1/3 of all points to be outside of the 1σ range. Second, the resampling process tends to overestimate the cluster substructure. This arises because our smoothed surface brightness maps do contain some residual noise due to Poisson statistics from the cluster and the background, and we then inject an additional component of shot noise when computing a fake cluster observations. Thus, the value of the statistic for mock observations may be biased, and the confidence intervals for the mock observations and the real ones are not expected to coincide. However, we expect that the true surface brightness and the inferred one would produce the samples of statistic values with similar variances. Therefore, we can use the variability of the simulated sample to determine the size of the error bars, but should center the error bars on the statistic value obtained for the real observation instead of the mean of the sample. A similar method of calculating uncertainties from simulated observations was used by Böhringer et al. (2010).

The method described above provides robust uncertainty estimates, but requires complicated machinery that generates adaptively smoothed maps and mock observations. We have used this machinery to perform substructure sensitivity tests, but one may want to use simpler uncertainty estimation methods when only interested in the uncertainty of asymmetry for a given observation. Therefore, we developed a simplified uncertainty estimation method which does not use the adaptive smoothing algorithm. We used subsampling method to determine the scatter in the measured asymmetry values. We generated mock observations that take a random half of the counts from the original observation and computed substructure statistics from them. The scatter in the resulting asymmetry values is expected to $\sqrt{2}$

larger than what we would obtained for the full sample, so we need to reduce these error bars by $\sqrt{2}$. This method avoids additional assumptions about clusters introduced by the adaptive smoothing algorithm, and is significantly simpler in implementation. We compared the error bars produced with both methods (Fig. 3.2), and found them to be similar.

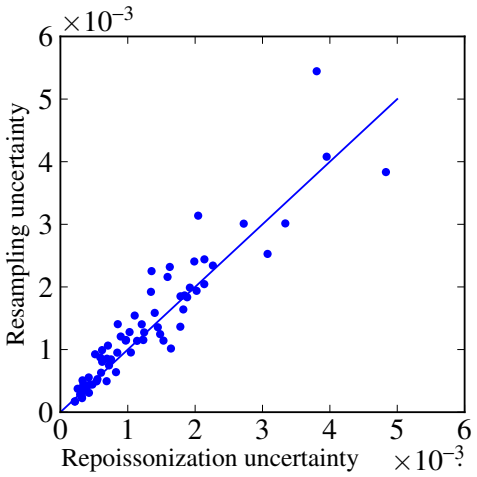

Figure 3.2: Comparison of A_{phot} uncertainties computed by two different methods. The horizontal axis represents uncertainties estimated by "repoissonization" (Weißmann et al., 2013b). The vertical axis represents uncertainties estimated by resampling half of the observation photons with replacement. The two methods agree well, suggesting that the simpler of the two (resampling) is a sufficient representation of the "true" uncertainty.

We also produced samples of 100 mock observations each where we changed one parameter of observation (such as exposure) for our sensitivity tests. In these tests, we viewed the adaptively smoothed images of clusters as the true surface brightness distributions in the sky. Unlike the previous group of simulations, here the true value of the statistic is not relevant. The median and the 68% confidence interval for each such sample represent how the statistic reacts to the corresponding change in the parameter of observation (such as exposure).

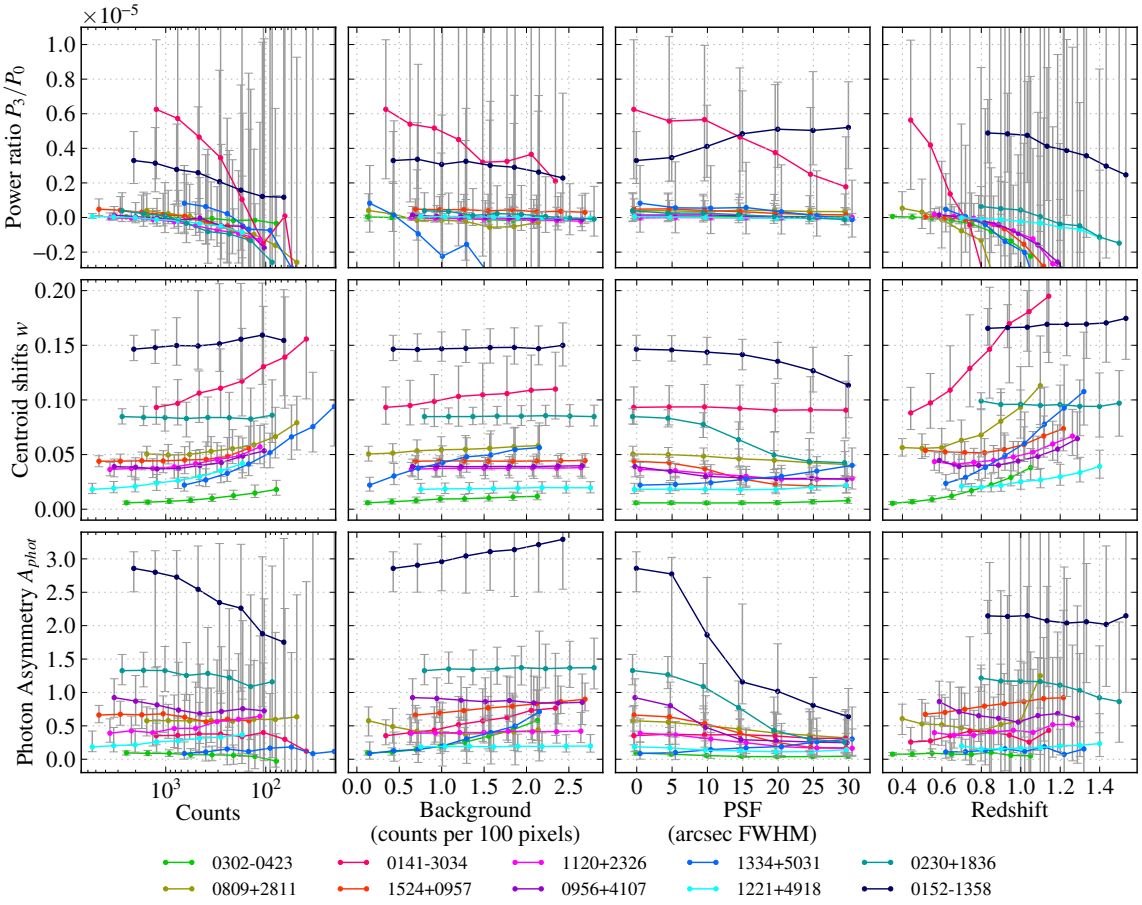

Figure 3.3: Sensitivity of 3 substructure statistics: power ratios, P_3/P_0 (1st row), centroid shifts, w (2nd row), and photon asymmetry, A_{phot} (3rd row) to the quality of observations: number of counts within R_{500} (1st column), background level (2nd column), telescope PSF (3rd column), and cluster redshift at fixed exposure (4th column). Only a representative subset of the clusters is shown in this Figure - the complete sample is shown in Fig. 3.4. An idealized morphological statistic would be insensitive to the quality of observations, i.e. all the lines should be parallel to x-axis. The photon asymmetry parameter presented in this paper shows better stability and better resolving power for observations of poor quality than commonly-used power ratios and centroid shifts. The names of the clusters are identical to those used in Vikhlinin et al. (2009a).

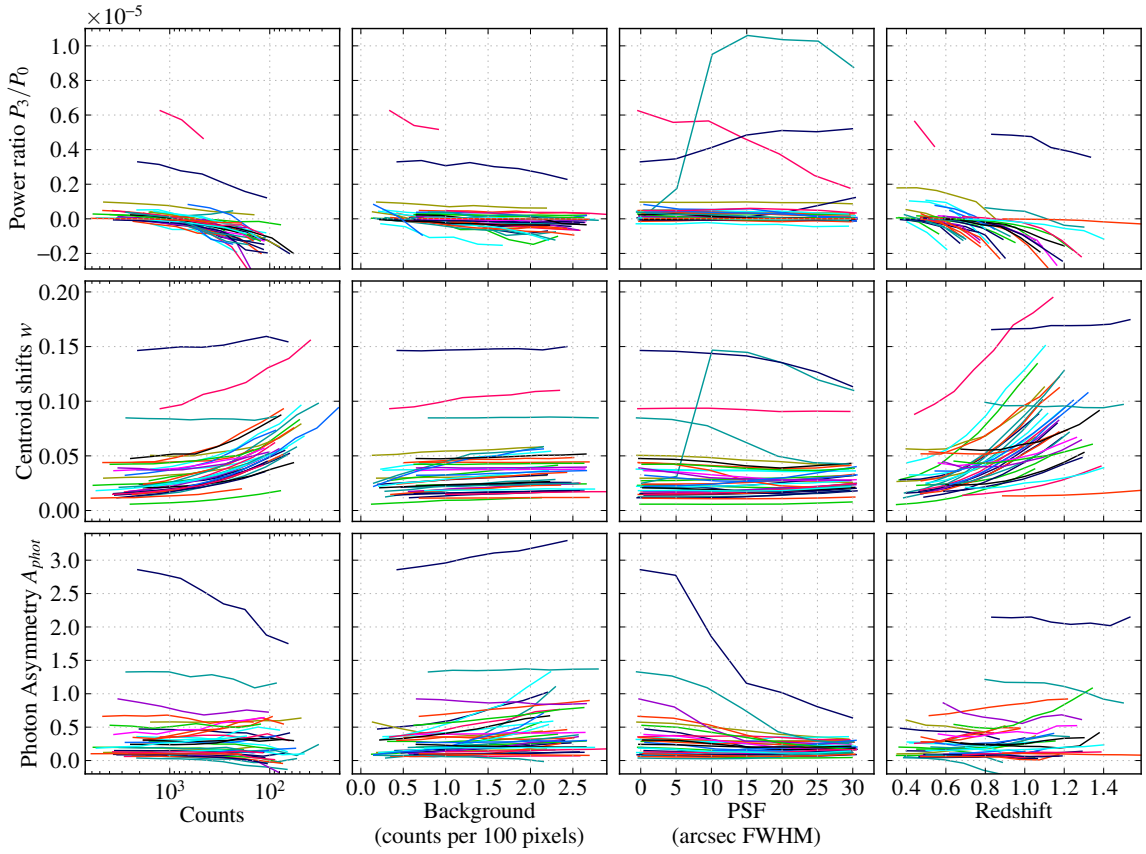

Figure 3.4: Similar to Fig. 3.3, but now showing the full sample of 36 clusters from Vikhlinin et al. (2009a). Uncertainties are excluded in this plot in order to reduce clutter – see Fig. 3.3 for a subset of these clusters with uncertainties. The points where uncertainties exceed the entire dynamic range for the corresponding substructure statistic are excluded from the plots.

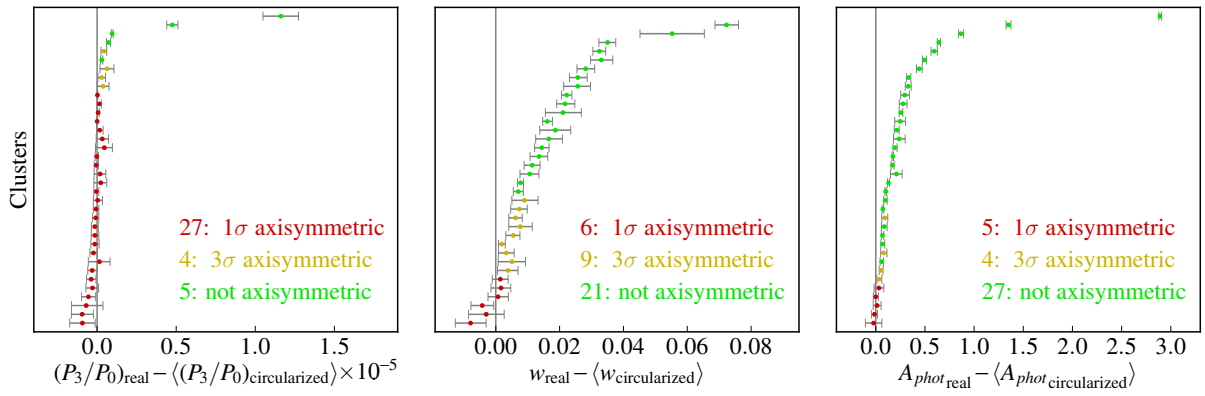

Figure 3.5: This figure demonstrates how well each of the three substructure statistics examined in this paper is able to discriminate between the true photon distribution and the idealized case of axisymmetry for each of the 36 clusters in Vikhlinin et al. (2009a). For each cluster, we compare the measured substructure statistic to the expectation for a perfectly smooth, axisymmetric source. Horizontal bars represent 1σ confidence intervals. The red points indicate clusters that are within 1σ of axisymmetry and, thus, are indistinguishable from the unrealistic case of perfect axisymmetry. The yellow and green points indicate clusters that are within 3σ and $>3\sigma$ of axisymmetry, respectively. Since no cluster should be *perfectly* symmetric, this plot demonstrates the sensitivity of each statistic to low levels of substructure, with $\rm A_{phot}$ performing the best overall.

3.6 Results and Discussion

3.6.1 Sensitivity of morphological parameters to data quality

An important test for any substructure statistic is its insensitivity to the observational S/N. Here we present sensitivity tests of two currently-popular substructure parameters (centroid shifts and power ratios P_3/P_0) and the new one introduced in this paper (photon asymmetry, A_{phot}). We conducted 4 tests that degraded the observations in different ways, namely 1) reduced the number of photons (exposure), 2) increased the level of background, 3) "blurred" the observation with larger PSF (or, alternatively, decreased the cluster's angular size), and 4) altered the observations in all mentioned ways, simulating an observation of the same cluster with the same exposure as if it was at higher redshift.

The plots of all sensitivities are presented in Figures 3.3 and 3.4, with different statistics in rows and sensitivity tests in columns. Fig. 3.3 shows 1σ confidence intervals, but only for a subset of representative clusters, while Fig. 3.4 shows median values of statistics that we obtained in our Monte-Carlo simulations (see Sec. 3.5.2). We chose to present plots of only P_3/P_0 for power ratios, because P_3/P_0 is believed to be the best indicator of substructure. The plots of P_2/P_0 and P_4/P_0 look qualitatively very similar.

All statistics show relative insensitivity to the number of cluster counts, at levels above $\sim$ 2000 counts. However, in the low counts regime, both power ratios and centroid shifts show strong biases. The power ratio tends to be biased low for all clusters. Centroid shifts tend to be biased high, more so for clusters that do not show significant substructure. Each cluster seems to have its own threshold value in number of counts, so that centroid shifts are stable when there are sufficient cluster counts, but start to increase as the simulated

number of counts falls below this threshold value. This behavior of centroid shifts is not surprising, because the statistical error of finding a centroid of a few points should scale as one over the square root of the number of photons, unless there are significant secondary emission peaks that "pin" centroids of certain radii. In other words, although this bias has a similar behavior for many clusters, it cannot be corrected simply as a function of number of counts – it also depends on the morphology (Weißmann et al., 2013b).

Centroid shifts, perhaps unsurprisingly, are the most stable statistic with respect to background levels. The determination of centroid is simply insensitive to a uniform background (unless there are so few counts that the $1/\sqrt{N}$ effect described in the previous paragraph starts playing a role). Power ratios are relatively stable with respect to background levels. (Although they become consistent with zero for *every* cluster in the sample after even a moderate background increment due to increased uncertainty.) Asymmetry is insensitive to background levels as long as a reliable estimate of cluster counts is possible in each annulus. However, when the square root of the total counts becomes comparable to the cluster counts, the estimate of cluster counts may become close to zero (or even negative). This unphysical estimate of cluster counts, being in the denominator in Eq. (3.19), drives the statistic to high absolute values. This is a drawback of $\mathrm{A_{phot}}$, which could be fixed by a more careful separation of background and cluster counts.

None of the statistics are stable against PSF increase because at $30''$ PSF the substructure is completely washed out and undetectable by any method. Asymmetry has a stronger sensitivity to PSF, because it probes the non-uniformity of the photon distribution on all angular scales, starting from the lowest Fourier harmonics to the highest. Power ratio P_3/P_0, on the other hand, is only sensitive to the third Fourier harmonic. It is interesting that the

PSF has a much stronger influence on any substructure statistic than does the number of counts. This observation suggests that for substructure studies a telescope's angular resolution is more important than its effective area.

The redshift test is the most challenging: the luminosity distance increases very fast, and the K correction adds to the flux dimming, effectively making the high-z simulated observations dominated by the background. Fluctuations in background increase variability of centroid estimation, driving centroid shifts to higher values. (A similar effect is demonstrated by sensitivity to cluster counts.) The power ratio median "dives" down to negative values (again, similar to counts test). Additionally, power ratio uncertainties increase very quickly, which is the result of background correction (subtraction of two nearly equal terms in Eq. (3.6)). Photon asymmetry also suffers from background correction, but overall shows less sensitivity to simulated redshift than either power ratio or centroid shifts.

What sets photon asymmetry apart from power ratios and centroid shifts is much smaller relative uncertainties. Unlike power ratios and centroid shifts, photon asymmetry is typically further than one standard deviation away from zero. So, photon asymmetry is capable of separating relaxed and slightly unrelaxed cluster populations in the case of observations with even a few hundred X-ray counts. To demonstrate that photon asymmetry is better than its competitors at distinguishing the clusters that are inconsistent with axisymmetric sources in the low S/N regime, we calculated the number of clusters in our sample that are 1σ consistent with circularly symmetric sources.

In order to compare the statistical significance of three different substructure parameters (photon asymmetry, centroid shifts and power ratios), for each cluster we generated a set of idealized, axisymmetric clusters. This was done by retaining the exact radial location

for each of the N detected photons for each cluster, but with a random realization of polar angles for each photon's position. We then computed the relevant structure metric. For each cluster we subtracted the mean of the parameter values computed from the fake circular clusters from that obtained from the actual cluster. We also assigned the scatter in the fake measurements as the uncertainty, for each cluster. Figure 3.5 shows departure from the circular case, with photon asymmetry clearly achieving a more significant determination of cluster substructure. Confidence intervals overlapping with 0 (red points) mean that the cluster indistinguishable from the axisymmetric case. The yellow points indicate clusters that are within 3σ of axisymmetry.

The number of clusters that are statistically inconsistent (at 3σ) with the idealized, axisymmetric case, as determined by different substructure statistics, are as follows:

- Power ratio P_3/P_0: 5 (out of 36)
- Centroid shifts w: 21 (out of 36)
- Photon asymmetry $\mathrm{A_{phot}}$: 27 (out of 36)

In other words, photon asymmetry has the best resolving power to measure "disturbance" in our sample.

The tendency of centroid shifts to be biased high for low-counts observations makes it questionable whether it can provide any meaningful results for samples of clusters with nonuniform S/N. We tested how the properties of the entire 400d sample would change if every cluster were moved to a greater redshift. In Figure 3.6, we plot the distributions of w and $\mathrm{A_{phot}}$ for the entire set of simulated observations at the original redshifts (blue), at the redshift $z + 0.3$ (green), and at the redshift $z + 0.6$ (red). We can see from the Figure that although the scatter is greater for the cluster sample at a higher redshift, the peak of

the A_{phot} distribution doesn't shift. This observation confirms that we can safely compare the values of asymmetry for cluster observations of significantly different S/N and redshifts. In Figure 3.6, the situation is different for w: the peak in its distribution shifts significantly moving to higher redshift, creating the false impression that higher-redshift clusters are more disturbed than their lower-redshift counterparts.

Overall, photon asymmetry is more stable with respect to changes in number of counts, background and redshift, and has smaller uncertainty than both centroid shifts and power ratios.

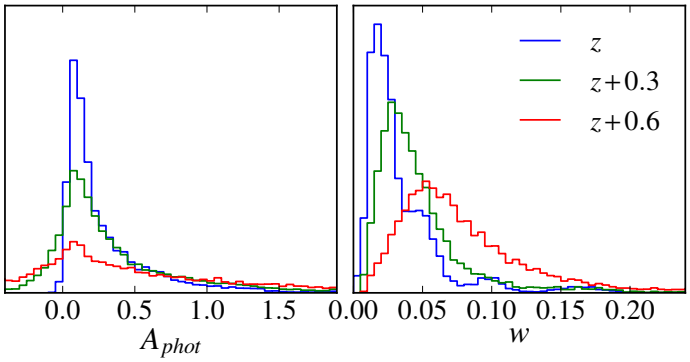

Figure 3.6: Distributions of substructure statistics for the entire sample of simulated observations at different redshifts. (Left) The peak of A_{phot} distribution doesn't move, for simulated observations at higher redshift. (Right) The peak of w distribution shifts to higher values as the clusters are shifted to higher redshift.

3.6.2 Asymmetry-concentration diagram

We propose a cluster classification scheme based on both concentration and asymmetry. Figure 3.7A shows the asymmetry-concentration diagram in logarithmic coordinates. The colors in Fig.3.7 are based on cluster "disturbance" as evaluated by-eye by a group of nine astronomers. Each participant was asked to score the disturbance of the clusters on the scale

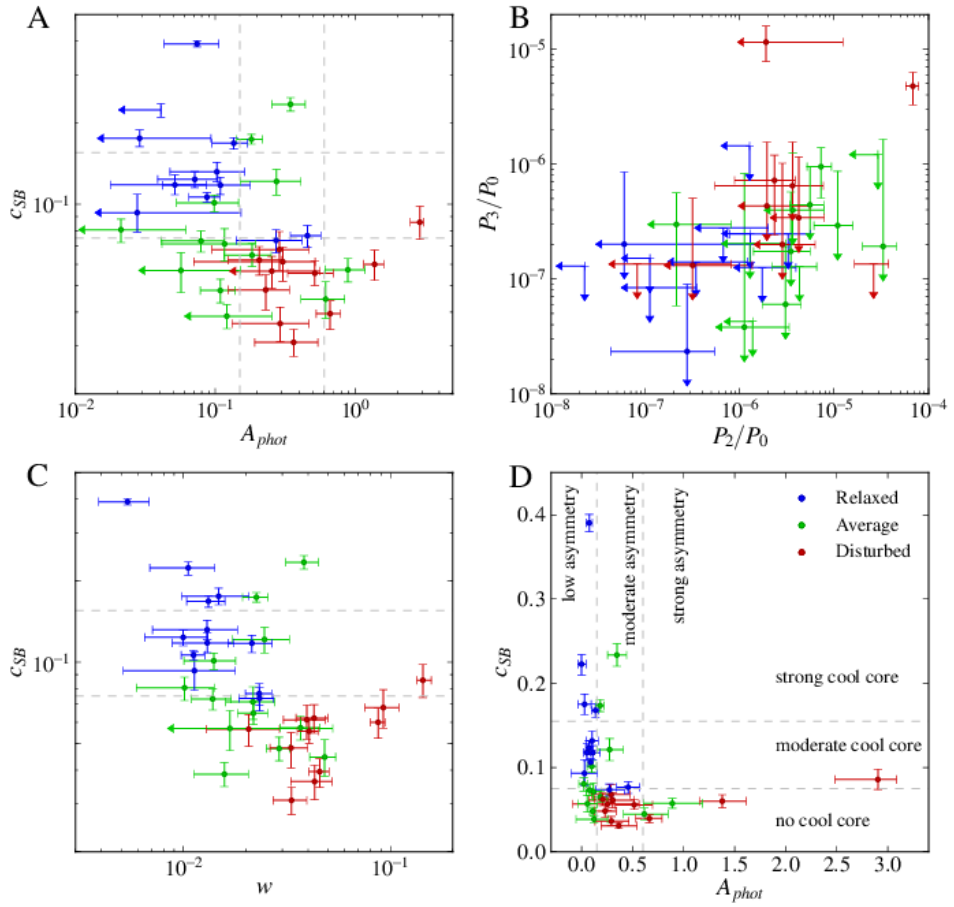

Figure 3.7: A: Cluster classification by A_{phot} (substructure statistic introduced in this chapter) and surface brightness concentration (c_{SB}; Santos et al., 2008). This classification scheme clearly separates relaxed, cool core clusters (high c_{SB}, low A_{phot}) from non-relaxed, disturbed systems (low c_{SB}, high A_{phot}). B: An alternative popular automatic classification scheme based on power ratios (Jeltema et al., 2005) – see text for details. Overall, the uncertainties are larger than in the asymmetry-concentration plane, and the clustering of same-type clusters is not as pronounced. C: Cluster classification using centroid shifts w instead of A_{phot}. Panel C achieves similarly good separation of relaxed and disturbed clusters, however the values of w may be correlated with observational S/N as shown in Section 3.6.1. D: Same as A, but in linear coordinates. This plot emphasizes that there are no clusters that are both highly concentrated and "asymmetric" The horizontal dashed lines splitting the range of concentrations correspond to cool core categories as defined by Santos et al. (2008). The vertical dashed lines splitting the range of asymmetries correspond to asymmetry categories (low, moderate, strong) as defined here In all panels, colors are assigned based on by-eye classification of "disturbance" (see text for details): blue - most relaxed, green - average, red - most disturbed.

1 to 3 (fractional values allowed), with 1 being least disturbed and 3 being most disturbed. We found that 11 of the clusters were unanimously ranked in the most disturbed half. We call this group of clusters "most disturbed", and mark them in red in Figures 3.7, 3.8, and 3.9. Another 12 clusters were unanimously placed in the least disturbed half of the rankings. We call this group of clusters "relaxed" and mark them in blue. The remaining 13 clusters are "average" and marked in green.

The asymmetry-concentration diagram (Fig. 3.7A) shows a significantly better separation of clusters at different states of dynamical equilibrium (as assessed by human experts) than the competing scheme of cluster classification based on power ratios proposed by Jeltema et al. (2005) and presented in Fig. 3.7B. The other drawbacks of the power ratios classification scheme are that $\log P_2/P_0$ and $\log P_3/P_0$ correlate (correlation coefficient = 0.61) and that both P_2/P_0 and P_3/P_0 are often consistent with 0. Therefore, what we see in Fig. 3.7B is mostly noise, whereas most clusters in Fig. 3.7A show a significant detection of substructure as discussed above.

A similar separation of clusters at different states of dynamical equilibrium can be achieved using w instead of $\mathrm{A_{phot}}$ as the substructure statistic (Fig. 3.7C), however $\mathrm{A_{phot}}$ is more stable and less biased for low S/N observations, as discussed above.

One can see that in Fig. 3.7A clusters avoid the upper right corner which confirms the standard assumption that concentrated or CC clusters are more regular. Fig. 3.7D, which is the same as Fig. 3.7A, but plotted in linear coordinates, shows this even better. It has a characteristic L-shape which implies that clusters are primarily either "concentrated" (upper part of the diagram) or "asymmetric" (right side) or "normal" (lower left corner).

In all the relevant panels of Fig. 3.7 we plot two dashed vertical lines as threshold values

that separate low-, medium-, and strong-asymmetry clusters. The threshold values are $A_{phot} > 0.15$ and $A_{phot} > 0.6$. The horizontal dashed lines separate strong, moderate, and no cool cores as defined by Santos et al. (2008). The threshold values are $c_{SB} > 0.075$, and $c_{SB} > 0.155$.

The asymmetry-concentration classification scheme makes a clear separation between the radial and the angular structure. Concentration only probes the radial photon distribution, while asymmetry probes the angular photon distribution. We expect these to be uncorrelated, a point to which the data attest (correlation coefficient = -0.20). We show how asymmetry compares with power ratios and centroid shifts in Fig. 3.8. A_{phot} and w are correlated strongly with correlation coefficient 0.87. This indicates that for high S/N data A_{phot} and w agree well on which clusters are disturbed.

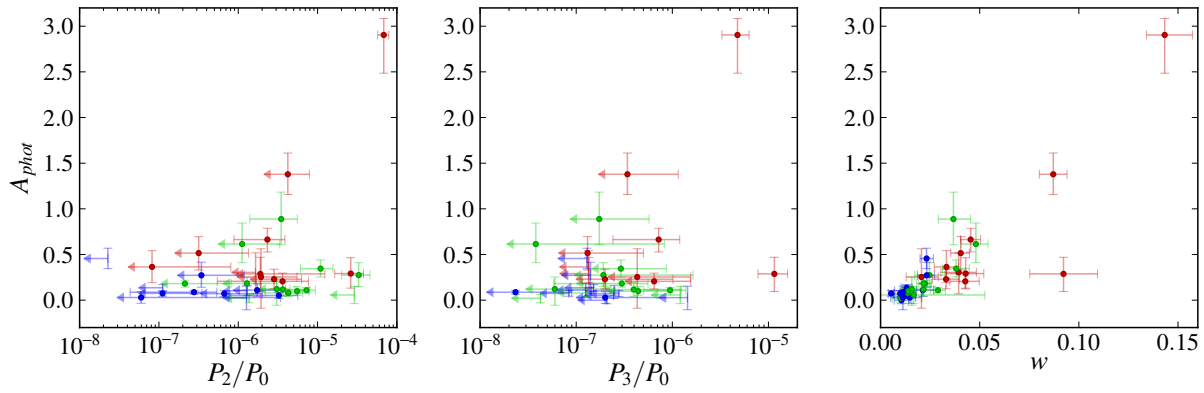

Figure 3.8: Comparison of photon asymmetry, A_{phot}, with power ratios, P_2/P_0 and P_3/P_0, and centroid shifts, w. Where power ratios are consistent with zero they are plotted with arrows as their lower limits. Where power ratios are negative, they are plotted with arrows at their upper limits. Power ratios with negative upper limits are absent from the plot. Colors are assigned based on by-eye classification of "disturbance": blue - most relaxed, green - average, red - most disturbed (see text for details). There is no obvious correlation between A_{phot} and the other substructure parameters, with the exception that they all tend to agree on the most disturbed systems.

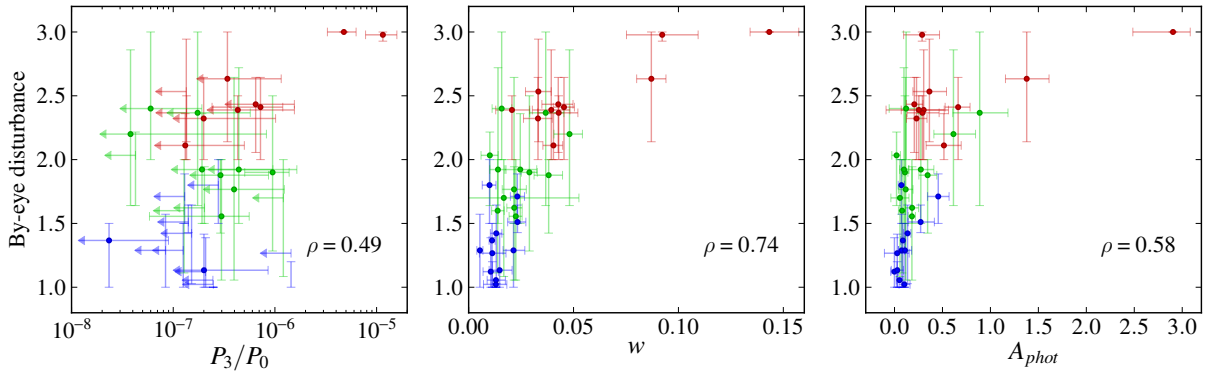

Figure 3.9: Comparison of 3 substructure statistics with the average by-eye disturbance score as evaluated by a group of nine astronomers. The correlation coefficient between the by-eye score and the corresponding statistic is shown on the plots. The colors based on by-eye disturbance score in the same way as in Fig. 3.7 and 3.8.

3.6.3 Relative ranking of clusters by the amount of substructure, by-eye classification

In Fig. 3.9 we show how photon asymmetry, centroid shifts and P_3/P_0 power ratio compare to by-eye classification. We find that the photon asymmetry parameter correlates with the human "by-eye" ranks almost as strongly as centroid shifts, with a Spearman's rank correlation coefficient of 0.71 for $\mathrm{A_{phot}}$, 0.75 for w. Power ratio P_3/P_0, on the other hand, shows much lower correlation coefficient 0.47.

In Figures 3.13, 3.11, 3.15, we present three side-by-side comparisons of morphological indicators. In each figure the left panel shows the (same) X-ray images of galaxy clusters, ordered by increasing values of our photon asymmetry parameter. The right panel shows these same clusters, ranked by increasing centroid shifts, power ratio P_2/P_0, and by-eye disturbance, respectively. To produce by-eye ranking, we averaged the disturbance scores (1 to 3) obtained from all nine human experts. We then ranked clusters by average disturbance

score.

3.7 Conclusions and future work

In this work, we introduced a new cluster substructure statistic – photon asymmetry (A_{phot}), that measures the uniformity of the angular X-ray photon distribution in radial annuli. We compared photon asymmetry to two other measures of cluster morphology, power ratios (with a novel method for background correction) and centroid shifts, on the 400d cluster sample, and on simulated observations derived from it. Our focus was on performance of these substructure statistics in the low S/N regime, that is typical for observations of distant clusters. Our main conclusions are as follows

1. The angular resolution of a cluster observation is far more important than total counts for the ability to detect and quantify the substructure.
2. Both centroid shifts and photon asymmetry are significantly more sensitive to the amount of substructure than power ratios.
3. Both centroid shifts and photon asymmetry agree well with by-eye classification.
4. Centroid shifts are the best-performing substructure statistic in the low spatial resolution ($\theta \gtrsim 5''$) and background-dominated ($\sqrt{F_{bg}} \gtrsim F_{source}$) regimes.
5. Photon asymmetry is the best-performing substructure statistic in the low-counts regime.
6. Photon asymmetry is the most sensitive measure of the *presence* of substructure; 27 out of 36 clusters in the sample are classified by photon asymmetry as clusters with

significant substructure (i.e., they are inconsistent with being axisymmetric), whereas the second best statistic, centroid shifts, finds significant substructure in only 21 out of 36 clusters.

7. Photon asymmetry is the only statistics that is insensitive to observational S/N below ~ 1000 counts. Consequently, it is the only statistic suitable for comparison of clusters and cluster samples across large range of S/N, counts, backgrounds and redshifts. It is the best candidate for studying the influence of substructure on bias and scatter in scaling relations.

We also suggested using concentration (a measure of cool core strength) and asymmetry (which quantifies merging or disturbance) as the main parameters for cluster classification. We find that clusters can demonstrate either a high degree of concentration or asymmetry, but not both at the same time. It is possible to use centroid shifts instead of photon asymmetry as the measure of cluster disturbance, but asymmetry is preferable given its better stability with respect to observational S/N.

We are currently applying the photon asymmetry metric in a comparison of X-ray and SZ-selected cluster samples, to study the impact of morphology on cluster scaling relations, and to measure how morphology evolves with redshift.

3.8 Acknowledgments

We thank Christine Jones-Forman, William Forman, Alexey Vikhlinin, Christopher Stubbs, Bradford Benson, Eric Miller, Michael McDonald, Marshall Bautz and Alastair Edge for their aid in morphologically classifying these galaxy clusters by eye.

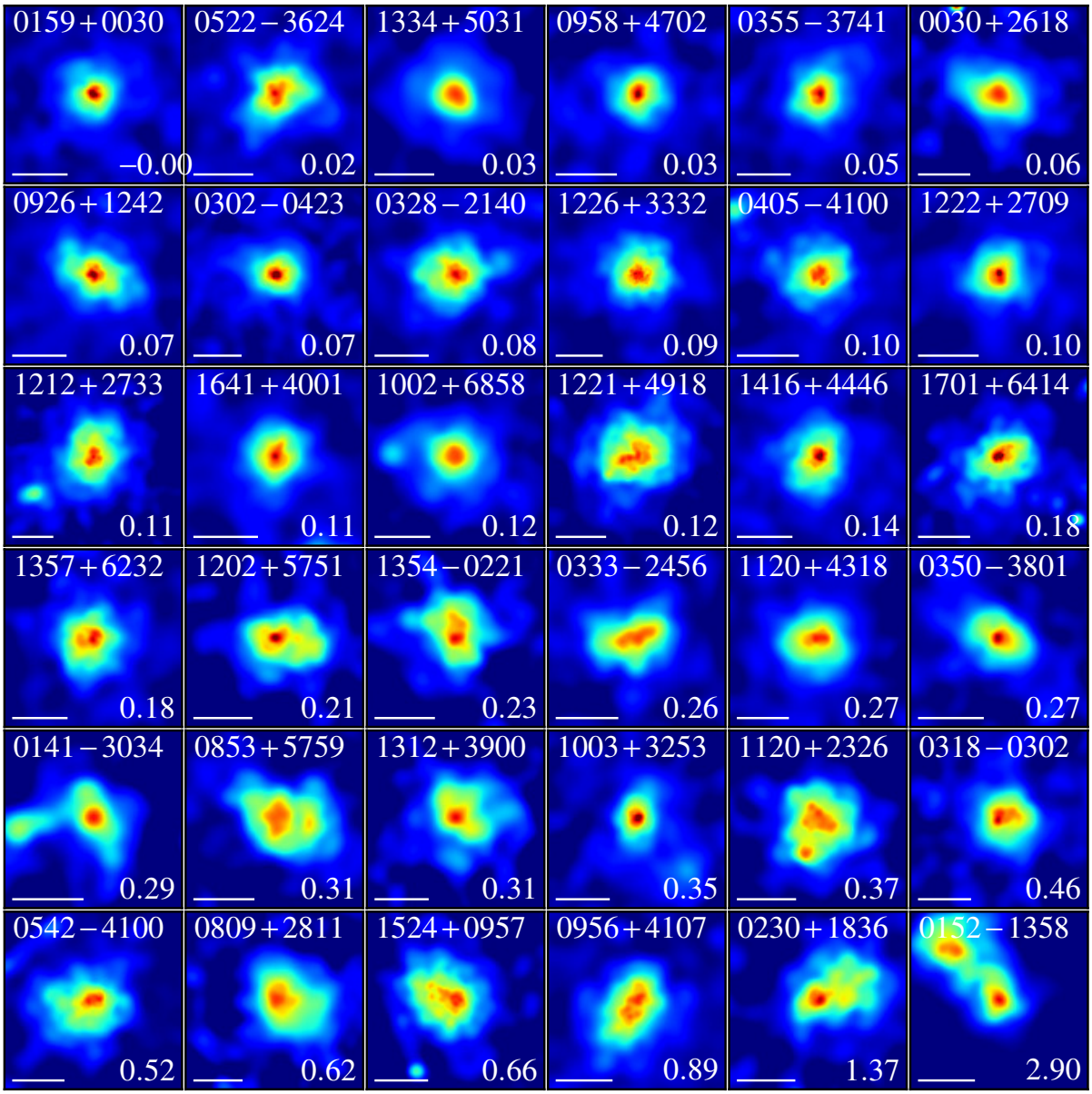

Figure 3.10: Clusters sorted by the value of asymmetry - the new substructure measure which is presented in this paper.

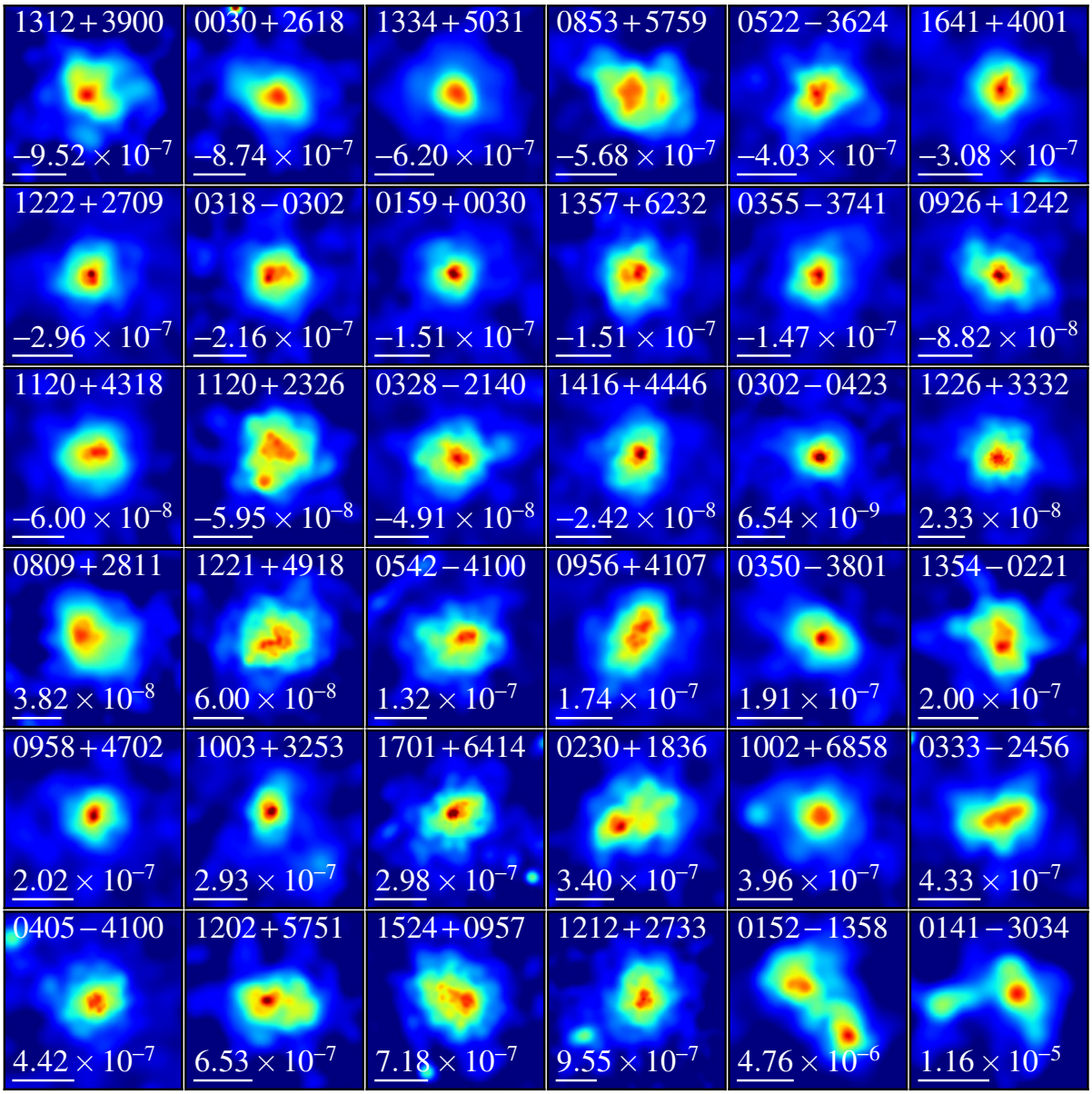

Figure 3.11: Clusters sorted by the value of P_3/P_0. White circle is the aperture used for calculating power ratio (R_{500}).

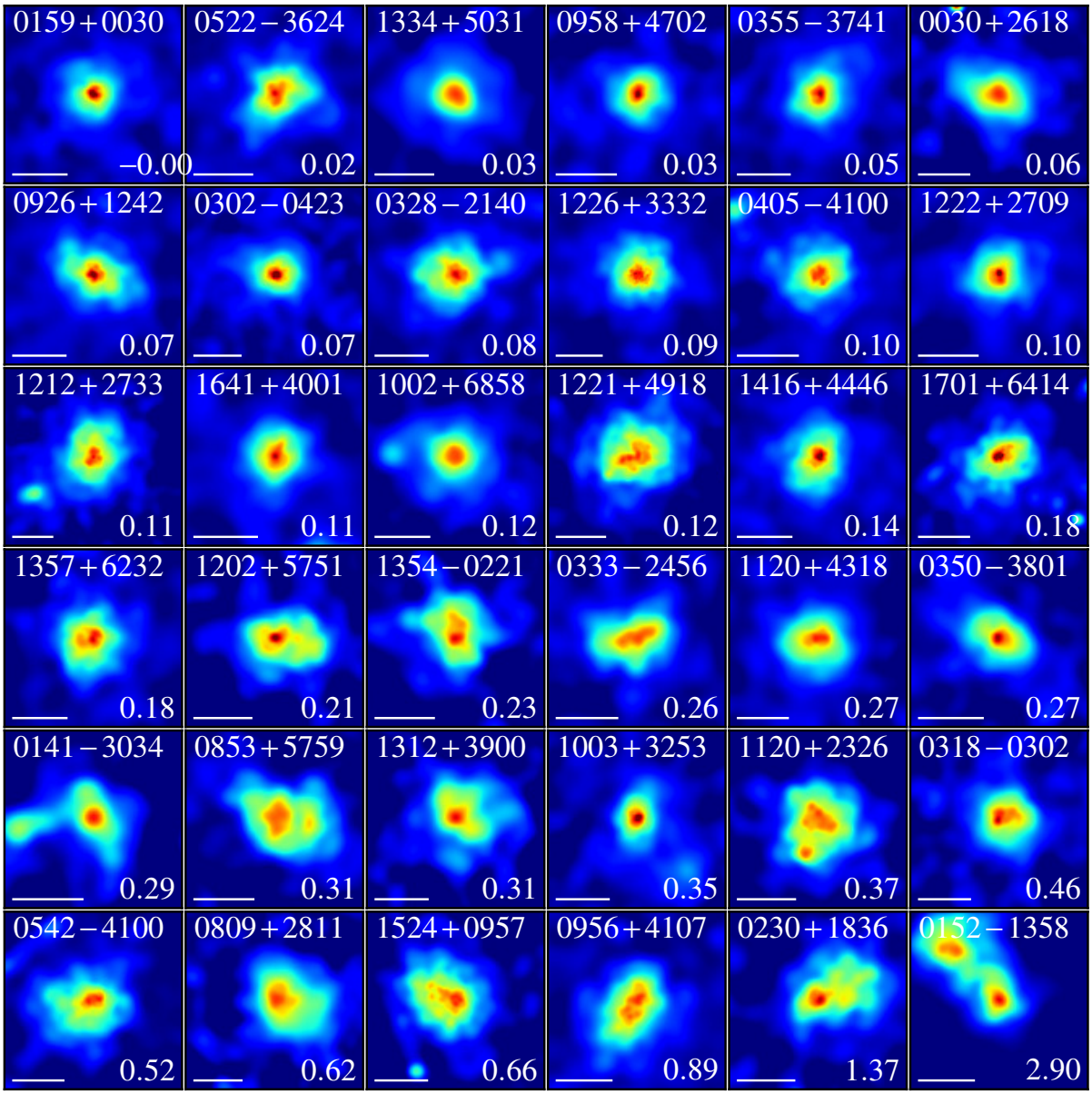

Figure 3.12: Clusters sorted by the value of asymmetry - the new substructure measure which is presented in this paper.

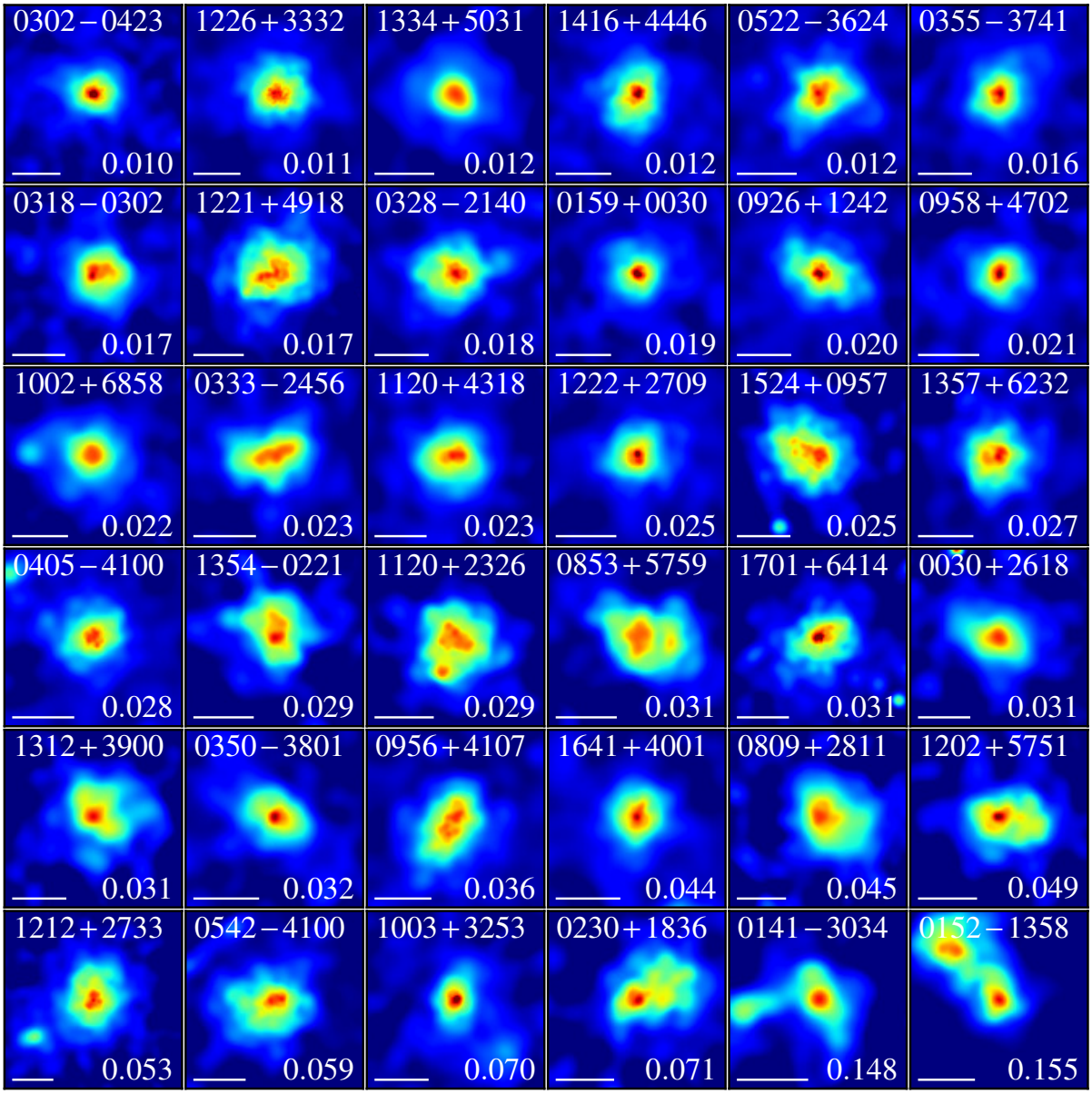

Figure 3.13: Clusters sorted by the value of centroid shifts.

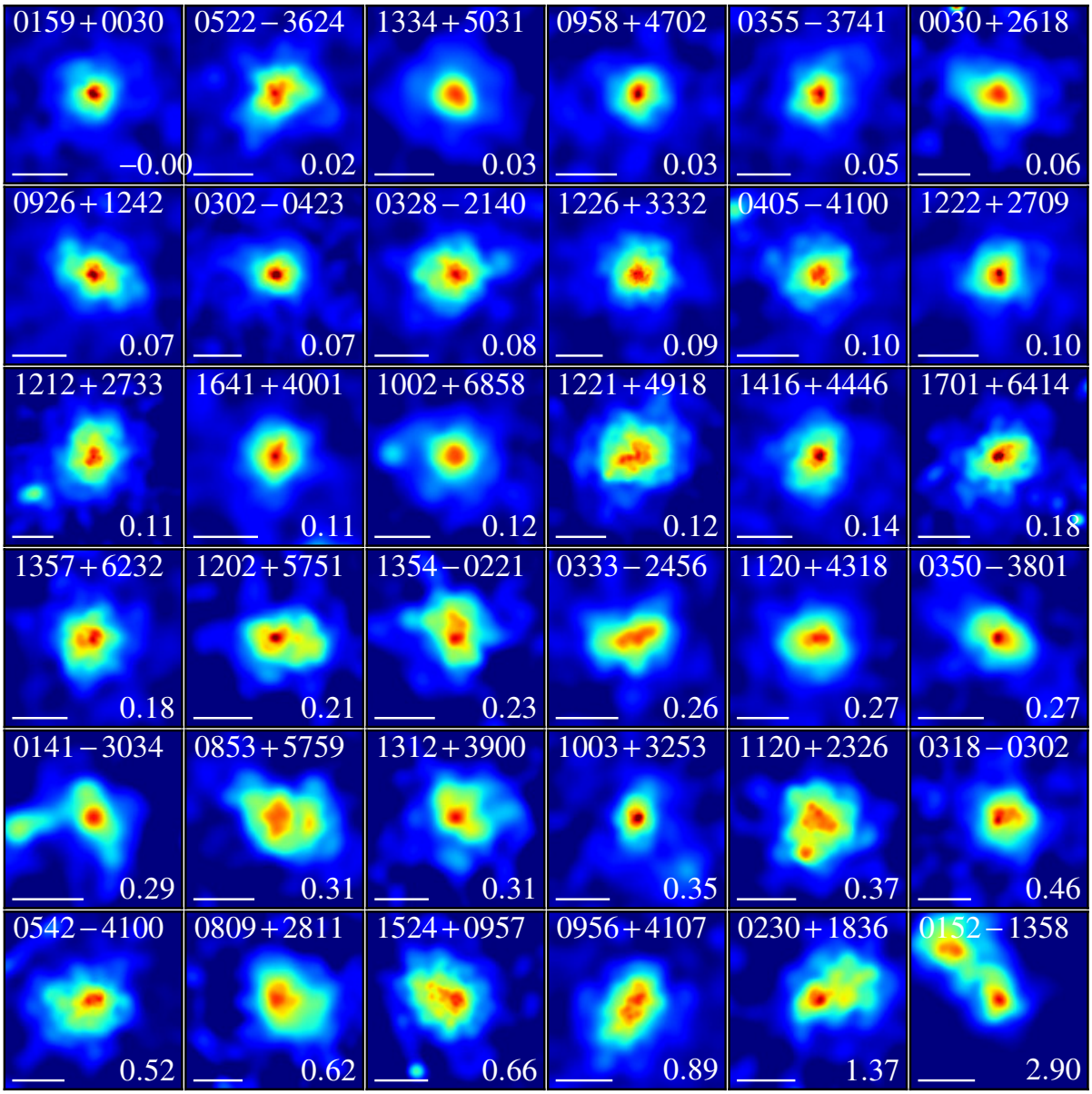

Figure 3.14: Clusters sorted by the value of asymmetry - the new substructure measure which is presented in this paper.

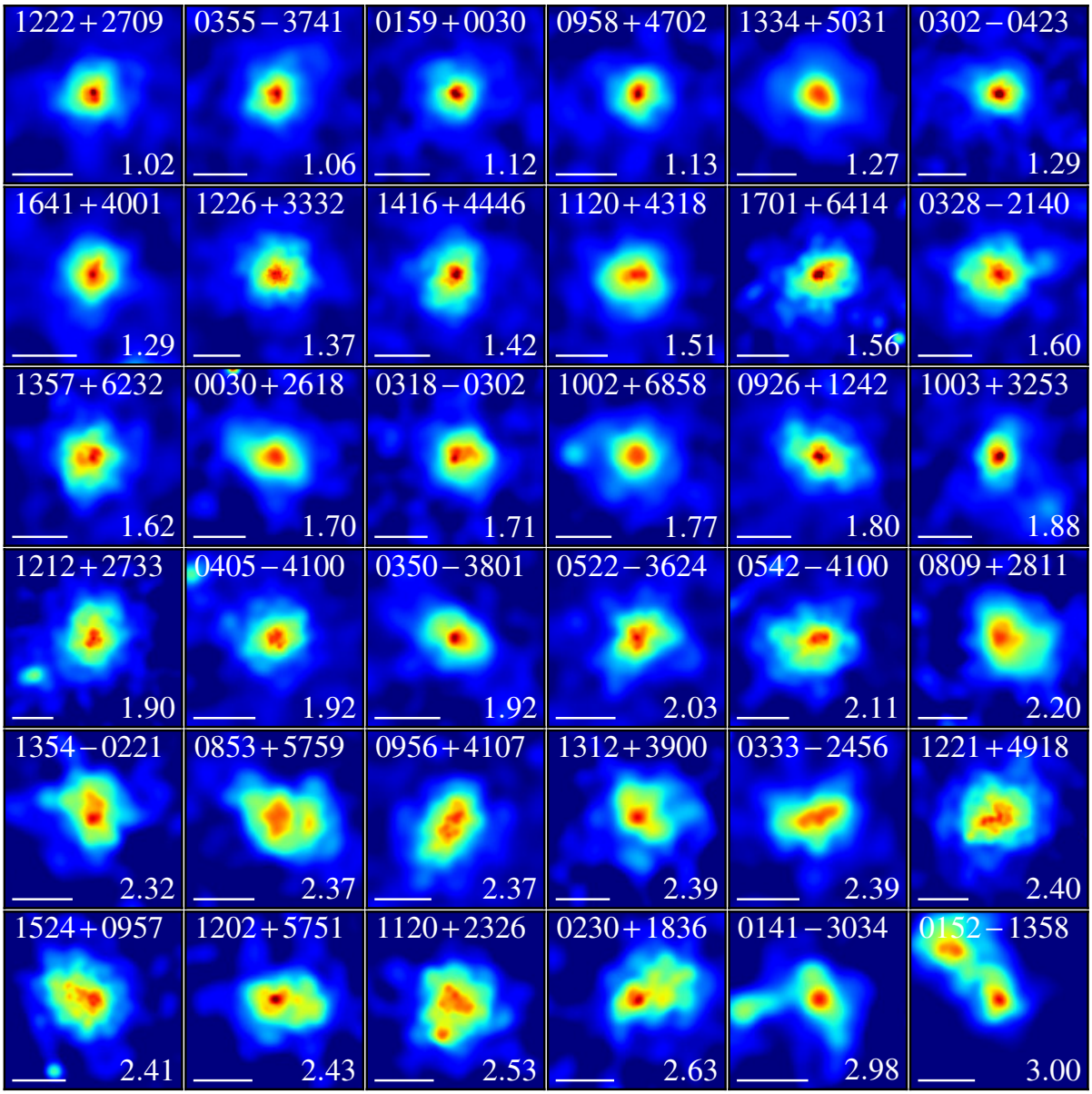

Figure 3.15: Clusters sorted by the average value of their "disturbness" evaluated by 4 numan experts.

3.9 Further details on asymmetry in the annulus

As explained in Section 3.3, our method of calculating asymmetry includes 2 steps: calculating the asymmetry in an annulus and combining the asymmetries from several annuli. To measure the asymmetry in each annulus we use the statistical framework of testing whether a given sample is drawn from a given probability distribution. The sample in our case is the empirical angular photon distribution function F_N, and the given probability distribution is the true angular photon distribution function G that would be produced by a perfectly circularly symmetric source. We note that G is not trivial because of nonuniform detector illumination and various detector imperfections.

We define F_N as the empirical cumulative angular distribution function of the photons in the k-th annulus:

$$F_N(x) = \frac{1}{N} \sum_{R_{in}^k < r_i < R_{out}^k} \mathbf{1}\{\phi_i/2\pi \leq x\}, \tag{3.23}$$

where $\mathbf{1}\{A\}$ is the indicator function of event A and N is the number of counts within the annulus $R_{in}^k < r < R_{out}^k$. Also, for convenience we rescale the angular range $[0, 2\pi)$ to $[0, 1)$. Let F be the true underlying distribution function for F_N, i.e F is the limit of F_N when $N \to \infty$.

Note that Kolmogorov-Smirnov, Cramer-von Mises and similar tests are usually used to check for the equality of 2 probability distributions. The values of these statistics give the probability of the null hypothesis (that the given sample is drawn from the given distribution), when compared to the null distribution. In our case, instead of checking whether F_N is a realization of the known F we need a measure of "distance" between F and G based on the measurement of F_N. In the following we show how one can use the value of Watson's test (a modification of Cramer-von Mises test suitable for distributions defined on a circle

as opposed to a segment) to quantify the distance between F and G based on the sample F_N.

In the following we will use the notation

$$U^2[F,G;dH] = \int \Big(F(x) - G(x) - \int \Big(F(x) - G(x)\Big)dH(x)\Big)^2 dH(x), \tag{3.24}$$

where F, G, and H are arbitrary distribution functions defined on $[0,1]$, and all the integrals are taken over the same $[0,1]$ interval.

Using this notation, Watson's statistic U_N^2 is simply

$$U_N^2 = N\, U^2[F_N, F; dF]. \tag{3.25}$$

It can be viewed as a minimum of the L_2 distance between F_N and F over all possible points of origin on the circle (Watson, 1961):

$$U_N^2 = \min_{\text{origin on the circle}} \int (F_N - F)^2 dF. \tag{3.26}$$

In the limiting case $N \to \infty$, under the null hypothesis that the sample ϕ_i comes from the hypothesized distribution $F(x)$, the values of statistic $U_\infty^2 = \lim_{N\to\infty} U_N^2$ have the same distribution as $K^2\pi^{-2}$, where K is distributed according to Kolmogorov's distribution:

$$\text{Prob}\{U_\infty^2 < x\} = \text{Prob}\{K < \pi\sqrt{x}\} = 1 - 2\sum_{k=1}^{\infty}(-1)^{k-1}e^{-2k^2\pi^2x^2}. \tag{3.27}$$

We won't need the exact form of this limiting distribution, but we need to know its mean which can be derived from known moments of Kolmogorov's distribution:

$$\langle U_\infty^2 \rangle = \frac{\langle K^2 \rangle}{\pi^2} = \frac{1}{12}. \tag{3.28}$$

It may be shown (Watson, 1961) that given a discrete sample $x_1, x_2, \cdots, x_N$ hypothetically distributed according to $F(x)$, the statistic can be computed as

$$U_N^2 = \frac{1}{12N} + \sum_{i=0}^{N-1}\left(\frac{2i+1}{2N} - F_i\right)^2 - N\left(\frac{1}{2} - \frac{1}{N}\sum_{i=0}^{N-1} F_i\right)^2, \tag{3.29}$$

where $F_i = F(x_i)$.

Now let's apply Watson's test statistic to the empirical distribution function F_N and an arbitrary distribution function G to which we need to compute a distance (in our method G is the distribution function that represents a circularly symmetric source)

$$W_N^2 = NU^2[F_N, G; dG]. \tag{3.30}$$

Integrating by parts one can show that

$$U^2[F_N, G; dG] = U^2[F_N, G; dF_N]. \tag{3.31}$$

Now we will replace dF_N with dF in the right hand side of (3.31). While it is evident (Doob, 1949) that as $N \to \infty$

$$U^2[F_N, G; dF_N] - U^2[F_N, G; dF] = R_N^2 \to 0 \text{ in probability,} \tag{3.32}$$

the merit of this approximation and the rate of convergence are discussed below.

$U^2[F_N, G; dF]$ can be transformed in the following way

$$\begin{gathered} U^2[F_N, G; dF] = \\ \int \left[\Big(F_N - F - \int (F_N - F) dF\Big) + \Big(F - G - \int (F - G) dF\Big) \right]^2 dF \\ = U^2[F_N, F; dF] + \\ \int \Big(F_N - F - \int (F_N - F) dF\Big) \cdot \Big(F - G - \int (F - G) dF\Big) dF + \\ U^2[F, G; dF] \end{gathered} \tag{3.33}$$

The first term, $U^2[F_N, F; dF]$ is distributed according to Kolmogorov's distribution and its mean is $\frac{1}{12}$ (see Eq. (3.28)).

The second term,

$$V = \int (F_N - F - \Delta) g(x) dx, \;\; g(x) = (F - G - \int (F - G) dF) F'(x) \tag{3.34}$$

has zero mean, because it is a sum of integrals of a function which has zero expectation value at any point on the segment

$$\langle F_N(x) - F(x) \rangle = 0 \;\; \forall x : 0 < x < 1 \tag{3.35}$$

with bounded functions $g(x)$ and $\int g(x) dx = const.$

The third term is the desired distance d between F and G.

Combining (3.30), (3.32) and (3.33) we find the following estimator $\hat{d}_N$ of d

$$\hat{d}_N = \frac{W_N^2}{N} - \frac{1}{12N}. \tag{3.36}$$

This estimator is biased by the average value of R_N^2.

$$\langle \hat{d}_N - d \rangle = \langle U^2[F_N, G; dF_N] - U^2[F_N, G; dF] \rangle = \langle R_N^2 \rangle \tag{3.37}$$

We were not able to obtain an analytic bound on R_N^2 and its N-dependence. Judging by the form of (3.32), R_N^2 should be of order $1/\sqrt{N}$. Considering this asymptotic behavior of R_N^2, our wish to explicitly correct for "smaller" bias $1/12N$ may look strange. The reason for this explicit correction is that $1/12N$ *is bigger* than R_N^2 for relevant values of N ($N < 10^3$). We confirmed this statement by multiple numerical experiments with various distribution functions F and G. As N reaches higher values ($N \sim 10^4$) $\langle R_N^2 \rangle$ can become greater than $1/12N$, but both terms tend to zero with increasing N.

Now we need to take into account that the acquired light comes both from the cluster and the background. We model the counts distribution function F as a weighted sum of

cluster emission F_{Cl} and a uniform background G

$$F = \alpha F_{Cl} + \beta G, \quad \alpha + \beta = 1, \quad \alpha = C/N, \tag{3.38}$$

where C is the number of cluster counts, and N is the total number of counts in the given annulus. Then we obtain

$$d = \int \Big(F - G - \int (F-G)dG\Big)^2 dG = \alpha^2 \int \Big(F_{Cl} - G - \int (F_{Cl} - G)dG\Big)^2 dG = \alpha^2 d_{Cl} \tag{3.39}$$

Now, using (3.36) we see that

$$d_{N,Cl} = \frac{N}{C^2}\Big(W_N^2 - \frac{1}{12}\Big) \tag{3.40}$$

is our estimator of the distance between the observed photon distribution function F_N and the underlying cluster emission distribution function F_{Cl}.

The sum of distances $d_{Cl}^{(k)}$ in 4 annuli, where k numbers the annuli, weighted by the estimated number of cluster counts in these annuli C_k, and multiplied by 100 gives *photon asymmetry*:

$$A_{phot} = 100 \sum_{k=1}^{4} C_k d_{N_k,Cl}^{(k)} / \sum_{k=1}^{4} C_k. \tag{3.41}$$

Chapter 4

Testing for morphological bias between X-ray and SZ-detected galaxy clusters

We present a study of galaxy cluster morphological properties as manifested in their X-ray emission as a function of their detection method and redshift. We analyze clusters from the 400SD catalog, which are X-ray luminosity selected, and from the XVP catalog, which are selected by their Sunyaev-Zeldovich signature. Clusters from both catalogs have similar-quality Chandra observations, which we use to characterize their X-ray morphologies. To quantify each cluster's morphology, we use a newly developed photon asymmetry (A_{phot}) parameter (Nurgaliev et al., 2013) and a more conventional centroid shift parameter. We do not find a significant difference between the morphological properties of X-ray and SZ-selected clusters. Also, we find that the morphologies of low-z and high-z subsamples are consistent with no evolution which suggests that the merging rate does not change

significantly between $z \sim 0.4$ and $z \sim 0.8$.

4.1 Introduction

Large scale galaxy cluster surveys can make an important contribution in understanding of the growth of structure in the Universe, delivering precise constraints on the nature of the dark energy, and providing insights into astrophysical processes in clusters. The primary interest in studying galaxy clusters from the cosmological point of view is in measuring their abundance as a function of mass and redshift. The abundance of galaxy clusters currently provides important constraints on cosmological models and parameters. There are many subtleties in interpreting abundance information from cluster surveys. First, as the mass is not directly observable, one needs to use scaling relations to link a cluster observable (such as X-ray spectroscopic temperature or integrated SZ signal) to its mass. Both scatter and potential biases in the given scaling relation have an effect on constraints when fitting cosmological models. Second, one needs to understand the survey's completeness and purity. Finally, more subtle selection effects such as increased sensitivity to a particular sub-class of clusters may play a role. An example of such a bias would be an increased sensitivity of X-ray flux-limited catalogs to cool core clusters: as cool core systems have higher luminosity than non-cool core systems of the same mass, a flux-limited catalog can potentially be biased towards cool core clusters (Hudson et al., 2010; Mittal et al., 2011; Eckert et al., 2011). As the ratio of cool core to non-cool core systems at different redshifts is not known and is a subject of active research (Vikhlinin et al., 2007; Santos et al., 2010; Samuele et al., 2011; McDonald, 2011; Semler et al., 2012; McDonald et al., 2013), this bias effectively limits our understanding of completeness in the X-ray flux-limited catalogs.

The primary biases of various detection methods are well known. Optical detection is biased to clusters richness. X-ray luminosity is proportional to gas density squared, so X-ray detection is biased towards cool core systems that have high central densities. A significant amount of SZ signal originates from the cluster outskirts – a large volume filled with a low pressure gas undetectable by other methods. Consequently, SZ detection is biased towards the gas properties in the outskirts. Since both X-ray and SZ detection methods are based on the physical properties of the ICM, they may have some common biases (Maughan et al., 2012) that are completely different from the optical detection methods which are sensitive to a different component of galaxy clusters. The finer details of each detection method's sensitivity to specific cluster morphology or dynamical state are not well understood.

In this paper we will focus on potential morphological biases in X-ray vs. SZ selection. It is commonly believed that SZ-selected clusters have higher merger rates. This belief is largely based on 1) the presence of spectacular mergers among the first few clusters discovered by SZ effect such as El Gordo (Menanteau et al., 2012) and PLCKG214.6+37.0 dubbed as "triple system" (Planck Collaboration et al., 2013b); and 2) an extensive discussion of newly discovered mergers in the papers which originated from the Planck *XMM-Newton* follow-up program (Planck Collaboration et al., 2011, 2012, 2013c). The latter program targeted 51 cluster candidates and led to the confirmation of 43 candidates, 2 of them being triple systems, and 4 double systems. The 37 remaining objects had 1) lower X-ray luminosity than expected from scaling relations and 2) shallower density profiles than the mean density profiles of X-ray detected clusters. These two observations served as the main arguments for Planck's increased sensitivity for mergers. While systems discovered by Planck do have interesting morphological properties, there are several arguments why Planck's findings do

not necessarily indicate the inherent sensitivity of the SZ effect to merging clusters:

1. A few of the double and triple systems discovered by Planck are clusters overlapping in projection, rather than interacting systems (although they still belong to the same supercluster structure). The increased sensitivity of Planck to such multiple systems is unsurprising due to its large PSF and consequent inability to resolve multiple systems.
2. No quantitative examination of the dynamical state of XMM-Newton observed clusters has been done. They, contrary to Planck's claims, may be equally disturbed as their X-ray detected counterparts.
3. The lower central density and luminosity of clusters may be related to greater than previously thought intrinsic scatter in these parameters, or other factors than merging processes.
4. Finally, finding 37 previously undetected single systems that appear to have higher substructure rates may be not so surprising given that the full Planck catalog contains 861 confirmed entries (Planck Collaboration et al., 2013a).

If SZ detection is indeed biased towards merging systems, it again has implications for abundance-derived cosmological parameters: it is more probable to find a merging system of a certain large mass at a particular redshift than a virialized cluster of the same mass.

Another question that has been extensively discussed in the literature and that we can address in our study is whether cluster morphologies change with redshift. The motivation for these studies is the connection of merger rate (and consequently morphology) to the matter density Ω_m. The morphology-cosmology connection that was analytically developed by Richstone et al. (1992) and then confirmed in simulations by Evrard et al. (1993); Jing

et al. (1995) predicted that clusters in low Ω_m models are much more regular and spherically symmetric than those in $\Omega_m = 1$ models. Consequently, there were efforts to constrain Ω_m by finding the fraction of clusters with significant level of substructure as defined by various substructure statistics: Mohr et al. (1995) used centroid shifts, Buote & Tsai (1995) used power ratios and Schuecker et al. (2001) developed a set of several specialized tests. The program of constraining cosmological parameters through substrucutre rates had not been very successful due to the difficulties of defining the "significant level of substructure", insufficient number statistics of both observed and simulated clusters and possibly the discrepancy between the theoretical definition of a merger and what is measured by a substructure statistic (Buote & Tsai, 1995).

Nowadays cluster abundance provides much more precise constraints on Ω_m than those obtained by merger fraction studies. Nevertheless, the question of evolution of substructure in galaxy clusters is still popular. Most authors reported statistically significant evolution (Mann & Ebeling, 2012; Andersson et al., 2009; Jeltema et al., 2005); fewer authors, e.g. Weißmann et al. (2013a) arrived at the conclusion that clusters are consistent with no evolution. Weißmann et al. (2013a) performed a study of substructure evolution similar to ours. They used a slightly different cluster sample and substructure statistics, but arrived at similar results (See Sec. 4.5 for more details).

We use X-ray morphology to compare the dynamical state of X-ray and SZ selected clusters. The X-ray observations exist for most massive clusters discovered by SPT (see Sec. 2 for details). The morphology of the X-ray emission is indicative of the dynamical state of the cluster (Buote, 2002). Sharp central peaks in the X-ray emission are associated with cool cores and the long periods of isolation needed for cool cores to develop. Spiral

structures suggest that the cluster has recently undergone a minor off-axis merger. Complex morphologies imply that there is an on-going major merger between two or more systems.

Our primary objective in this paper is to test for any evidence of a morphological difference between X-ray and SZ-selected clusters. This is of particular interest if we wish to combine the X-ray and SZ samples in order to obtain better statistics of morphological studies.

4.2 The data

We used the high-z part of 400 square degrees cluster survey (Burenin et al., 2007) abbreviated here as 400d as our X-ray luminosity limited catalog. This high-z 400d catalog consists of 36 clusters in the redshift range $0.35 < z < 0.9$ and mass range $10^{14} M_{\odot} < M_{500} < 5 \times 10^{14} M_{\odot}$.

We used the 80 most significant clusters detected in the 2500 deg^2 SPT-SZ survey as our SZ-significance-limited catalog (Benson et al., 2014). X-ray images of these these clusters were obtained in the Chandra X-ray Visionary Project (PI B. Benson) abbreviated here as XVP. The XVP sample consists of 80 clusters spanning the redshift range $0.3 < z < 1.2$ and the mass range $2 \times 10^{14} M_{\odot} < M_{500} < 20 \times 10^{14} M_{\odot}$.

Both samples have similar quality Chandra observations. The exposures are typically sufficient to obtain ~ 2000 X-ray source counts. The high-resolution Chandra imaging with sufficient photon statistics is crucial to detect substructure in galaxy clusters (Nurgaliev et al., 2013). The 400d and XVP catalogs are currently the best available samples of high-redshift clusters with clear selection criteria and high-quality X-ray imaging. A good

overlap in the redshift ranges gives us an opportunity to select subsamples with similar redshift distributions.

The X-ray data reduction steps for both samples are equivalent to those described in Vikhlinin et al. (2009); McDonald et al. (2013) and Benson et al. (2014) and include filtering for flares, applying exposure maps corrections, detection and masking of point sources.

4.3 Methods

We perform our morphological analysis based on 2 morphological parameters that trace the degree of cluster disturbance: photon asymmetry $A_{\rm phot}$ (Nurgaliev et al., 2013) and centroid shifts w (Mohr et al., 1993). $A_{\rm phot}$ and w show a significant degree of correlation with each other and with by-eye classification of a cluster's dynamical state. Both are sensitive to irregularities of X-ray emission in the plane of the sky. We believe that $A_{\rm phot}$ has more statistical power in resolving substructure and is able to produce more consistent results independent of the quality of observation (such as exposure, background, etc). However, w is currently more popular and is a well established substructure statistic used by many groups, so we include it in our tests as well.

To judge the similarity of two cluster subsamples we use the 2-sample KS test for the empirical distributions of measured substructure parameters. The value of the KS statistic is converted to the p-value of the null hypothesis, that these two empirical distributions come from the same underlying distribution. Common practice is to reject the null hypothesis for $p < 0.05$.

Fig. 4.1 shows the distribution of XVP and 400d clusters in the (z, M) plane. As one can

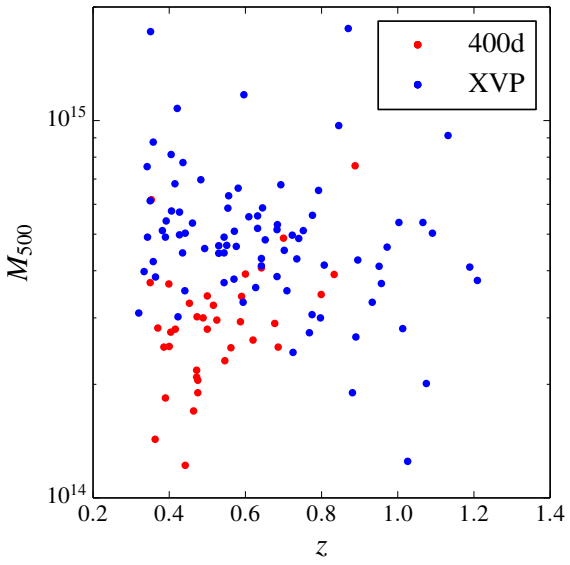

Figure 4.1: Masses and redshifts for XVP and 400d

see, there are proportionally fewer clusters beyond $z = 0.6$ in 400d sample. Therefore, for a fair comparison (which would be free of redshift evolution effects) we compare morphologies in the subsamples $z < 0.6$ of both catalogs.

We use the following pairs of subsamples in our comparisons:

1. Full 400d $z < 0.6$ and XVP $z < 0.6$. These can be viewed as maximally similar catalogs in redshift space. This is the cleanest test for the difference between X-ray and SZ selection functions.

2. XVP $z < 0.6$ and XVP $z \geq 0.6$ is the cleanest test for substructure evolution, because both subsamples come from the same catalog with a well-defined selection function.

3. 400d $z < 0.6$ and XVP $z \geq 0.6$ is a complementary test for substructure evolution. Although, 400d and XVP catalogs have very different selection functions, they have

very similar distributions of substructure statistics. Thus, this test is also meaningful.

4. Full 400d and full XVP. Do the morphologies still look the same in the complete SZ and X-ray selected catalogs regardless of the different redshift ranges that these methods pick up?

5. Combined 400d and XVP split into $z < 0.6$ and $z \geq 0.6$ parts. If selection criteria are indeed not important, this test increases the resolving power of the substructure evolution test.

4.4 Results

Fig. 4.2A shows histograms and cumulative distributions of A_{phot} and w for 400d and SPT-XVP $z < 0.6$ subcatalogs. High p-values of the KS statistics (0.89 and 0.57 respectively) indicate that the distributions of statistics values for these two data samples are indistinguishable. Under the assumption that mergers are characterized by increased values of A_{phot} and w , this means that the amount of merging systems detected by SPT is similar to the amount of merging systems detected by their X-ray emission.

As the next step we compare clusters at low redshift and high redshift. We consider $z < 0.6$ subsamples of 400d and XVP independently. Fig. 4.2B, C show histograms and cumulative distributions for the low-z part of 400d, and low- and high-z parts of XVP. According to A_{phot} , the high-z subsample is indistinguishable from both low-z subsamples. The w statistic shows a slightly lower p-value (but still far away from the 5% threshold), which is consistent with our claim that w may have a slight bias for high-redshift systems (Nurgaliev et al., 2013).

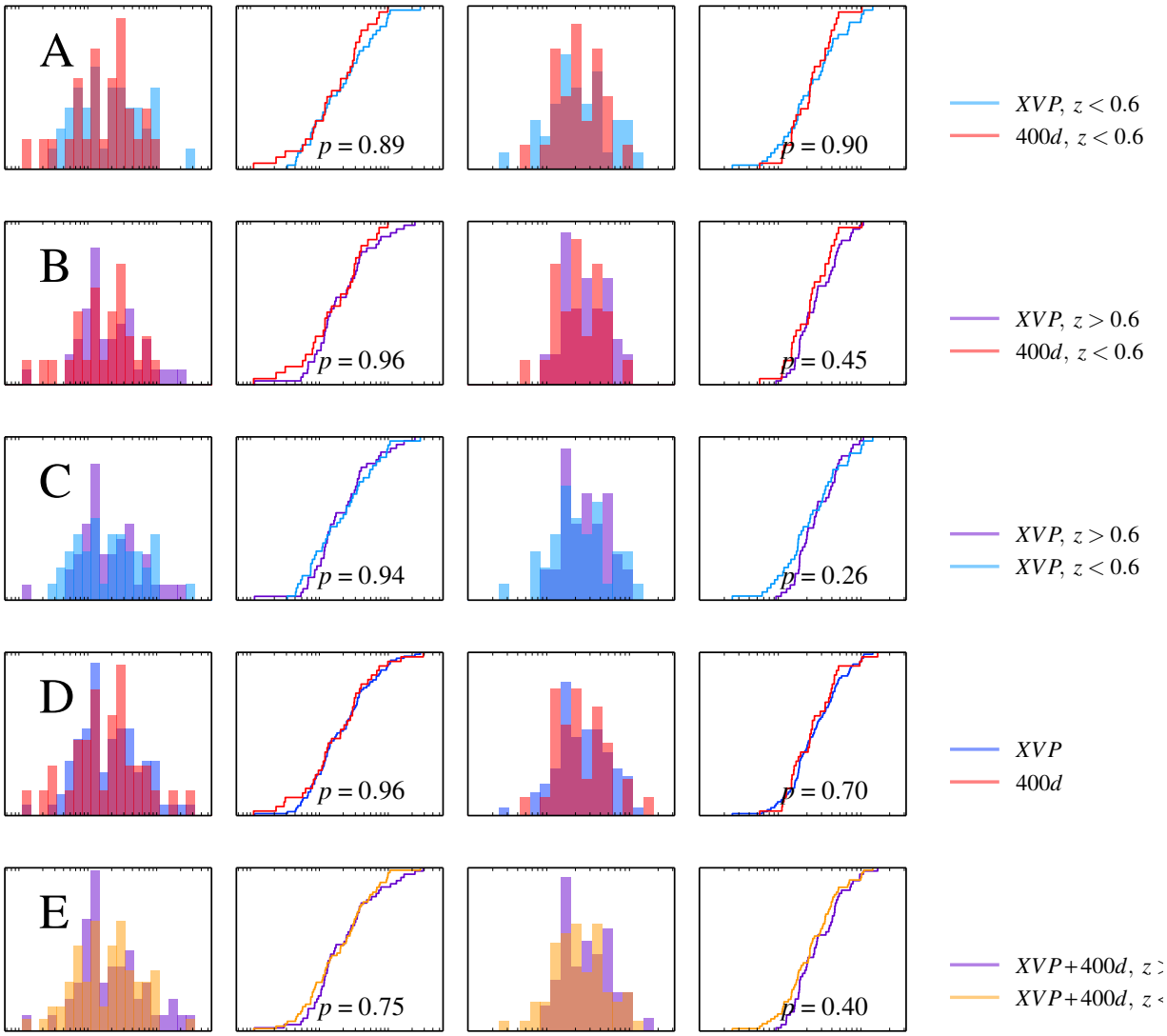

Figure 4.2: Distribution of morphological statistics (A_{phot} in the 2 left columns and w in the 2 right columns, histograms and cumulative distributions) for various partitions of clusters in 400d and XVP catalogs. From top to bottom: Full 400d $z < 0.6$ and XVP $z < 0.6$; XVP $z < 0.6$ and XVP $z \geq 0.6$; 400d $z < 0.6$ and XVP $z \geq 0.6$; Full 400d and full XVP; Combined 400d and XVP split into $z < 0.6$ and $z \geq 0.6$ parts. See also Sec. 4.3 for the motivation for these partitions and Sec. 4.4 for the description of the results.

Now we compare the full 400d catalog with the full XVP catalog. As we have already shown, our samples of clusters demonstrate very weak, if any, redshift evolution. Therefore, mixing clusters of slightly different redshift ranges shouldn't affect our comparison of the selection function's effects on observed amount of substructure in these catalogs. Fig. 4.2D shows distribution functions for the entire 400d and XVP catalogs. These distributions are still indistinguishable.

Finally, we compare $z < 0.6$ clusters with $z \geq 0.6$ clusters from both catalogs. This allows us improve the statistical power of the KS test under the assumption that the selection effects are indeed not important. Fig. 4.2E that there is no difference between these samples.

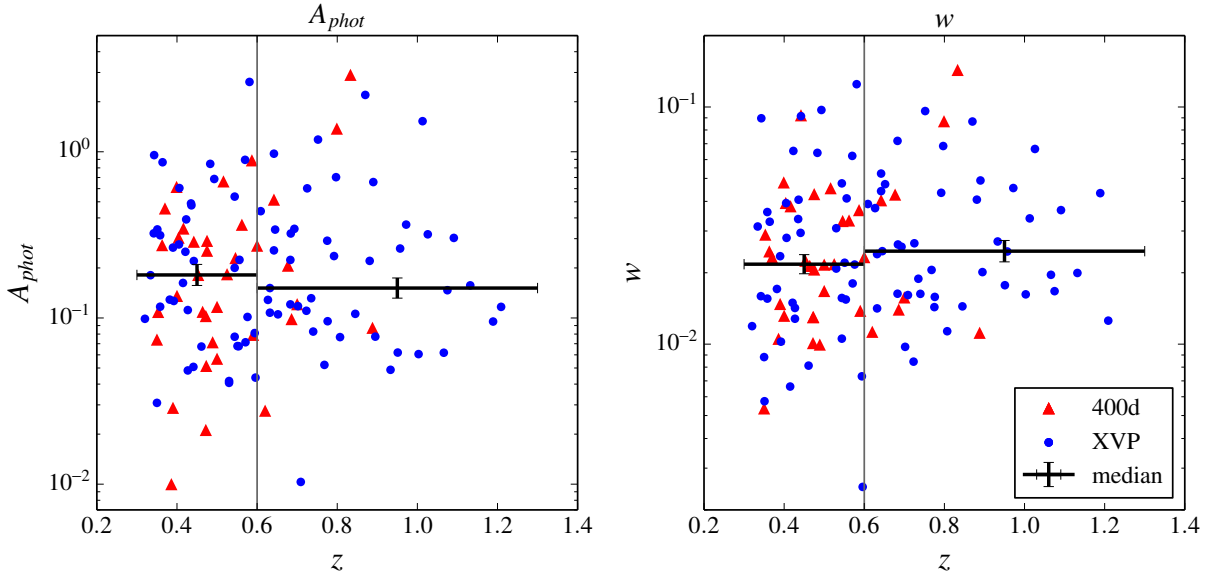

Figure 4.3: Redshift evolution of A_{phot} and w . Error bars on the median value of statistic is obtained by Median Absolute Devitation method and show that there is no significant difference in the median value between low-z and high-z subsamples.

Figure 4.3 shows the amount of substructure as a function of redshift. Although we see a group of more relaxed clusters for lower z subsample, on average clusters show the same

amount of substructure independent of redshift.

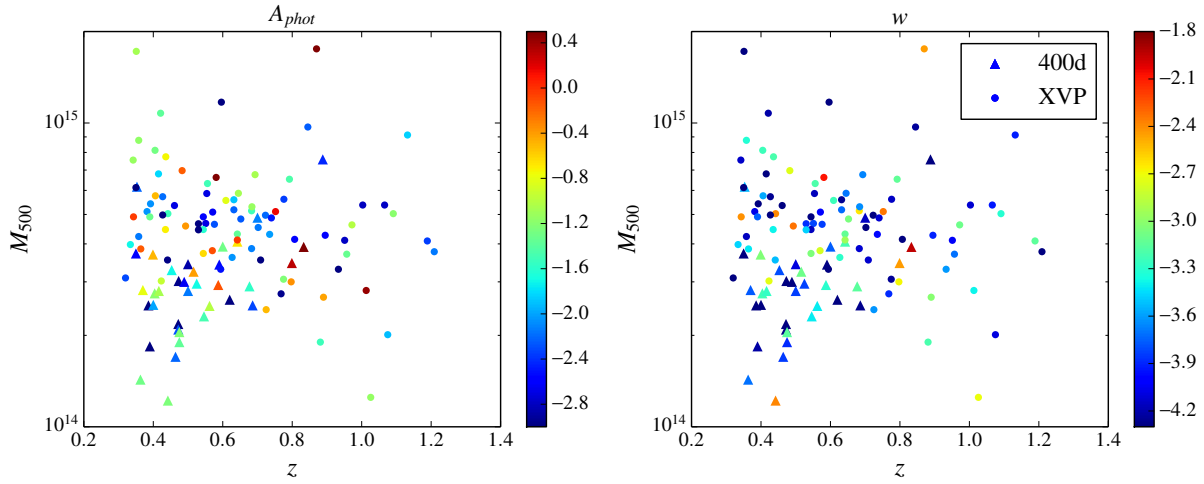

Figure 4.4: A_{phot} and w as functions of mass and redshift. There is no preferential aggregation of red or blue points in different parts of the diagram

Figure 4.4 is an alternative visual representation of our results: clusters on (z, M) plane are colored according to the value of their substructure score. There is no visible correlation of morphological parameters with mass, redshift or detection method.

4.5 Discussion

4.5.1 X-ray, SZ selection biases

We do not find any evidence of morphological bias between X-ray and SZ-selected samples. To reconcile our findings with Planck's results we'd like to reiterate 2 points mentioned in the introduction.

1. Due to extended big PSF, Plank does not resolve close galaxy clusters, e.g. PLCKG214.6+37.0 thereby capturing an "inflated" single SZ-detection where an instrument with smaller

PSF such as SPT would see several independent systems with smaller significance. Since the SPT beam is matched to the angular cluster size for $z > 0.3$, it is unlikely that SPT sample contains similar blended systems.

2. The detection of a few dozen previously unknown mergers compared to more than 800 confirmed objects in Planck's catalog is not a statistically significant indication of morphological bias in SZ-selected cluster samples.

4.5.2 Evolution of substructure with redshift

Our result that low- and high-redshift systems show the same amount of substructure is somewhat surprising – in the standard growth of structure scenarios the fraction of disturbed clusters increases with the redshift. Additionally, our findings of no redshift evolution are in contradiction with some earlier works, especially Jeltema et al. (2005) and Andersson et al. (2009). Both Nurgaliev et al. (2013) and Weißmann et al. (2013b) have found that shot noise has a very strong influence on the value of power ratios. Andersson et al. (2009) seemed to not apply any shot noise correction, so their results of a strong redshift evolution are not surprising.

One reason why we don't see an increased fraction of disturbed clusters may be that our low-z sample still has a relatively high median redshift of ~ 0.45. The median redshift of the high-z subsample is ~ 0.77.

Another reason is that there is not a one-to-one correspondence between substructure statistics such as w or P_3/P_0 and dynamical state of the cluster – as shown in simulations, substructures statistics may vary significantly on short time scales during cluster mergers (Hallman & Jeltema, 2011; Poole et al., 2006; O'Hara et al., 2006). For example, the O'Hara

et al. (2006) simulation results show that w can easily vary in the range $0.01 < w < 0.1$ over a fraction of characteristic merger time and Poole et al. (2006)'s simulations show even greater variations in centroid shifts in similar settings. In addition w shows similarly big fluctuations as a function of line of sight.

Finally, as pointed out by Weißmann et al. (2013a), earlier studies that used power ratios might not account correctly for the insufficient photon statistics of high-z clusters. For example, Jeltema et al. (2005) analyzed redshift evolution using a relatively small sample of 40 clusters divided into low and high redshift subsamples. The low redshift contained 26 clusters with $z < 0.5$ and $\langle z \rangle = 0.24$. The high redshift sample contained 14 clusters with $z > 0.5$ and $\langle z \rangle = 0.71$. They used the power ratio method and found the amount of substructure significantly different between the subsamples as measured by P_3/P_0. They also fitted the $P_3/P_0 - z$ relation and found the slope to be positive with high ($p \approx 0.005$) significance. Their results are surprising in the light of new studies of the properties of the power-ratios method. Both Nurgaliev et al. (2013) and Weißmann et al. (2013a) find that P_3/P_0 are consistent with zero for a majority of high-redshift clusters due to insufficient quality of observations.

Weißmann et al. (2013a) performed a similar study using the same high-z 400d sample and a subsample of XVP used in Andersson et al. (2011). In addition they used a low-z sample published in Weißmann et al. (2013b). They used the centroid shifts and the power ratio methods to quantify the substructures. They found that the amount of substructure seen in their samples is consistent with no redshift evolution. We would like to express caution against some of the techniques presented in Weißmann et al. (2013a): 1) use of upper limits of uncertainties in regressions for data points consistent with 0 (this might

drive the positive slope in $P_3/P_0 - z$ relations since P_3/P_0 for many high-redshift clusters is consistent with 0; and 2) using mean values of substructure statistics for low- and hig-redshift subsamples – because the distributions of substructure statistics are closer to lognormal than to normal, the mean is strongly affected by only a few objects with anomalously high substructures.

Weißmann et al. (2013a) put lots of emphasis on bringing both high-redshift and low-redshift subsamples to the same quality of observations. To achieve that, they artificially degrade higher-quality observations of the low-redshift sample. This is not necessary in our analysis because 1) the observations of both samples were targeted for 2000 counts per cluster and 2) as shown in Nurgaliev et al. (2013) w is not sensitive to number of counts above ~ 1000 counts (this is also confirmed in Weißmann et al. (2013b)) and $A_{\rm phot}$ has even better stability than w with respect to number of X-ray counts.

Although we do not find evidence for change of amount of substructure with redshift based on the robust non-parametric statistical test, one can speculate that there are systems in the lower redshift subsample with unusually low w and $A_{\rm phot}$, so our findings do not contradict earlier studies. However, based on the KS test and a robust comparison of the medians in different subsamples, we cannot call this a significant effect. Both subsamples could be drawn from the same underlying distribution.

4.6 Conclusions

We used a sample of 36 X-ray selected clusters and a sample of 80 SZ-selected clusters to test whether these selection methods have any preference towards cluster morphological

type, and to test whether high-redshift clusters are more disturbed than their low-redshift counterparts. We used the 2 most powerful substructure statistics (Nurgaliev et al., 2013), looking for differences in their distributions between different subsamples. We didn't find any morphological difference between X-ray and SZ-selected clusters, low-redshift or high-redshift.

Our results show that the amount of substructure we see in clusters from X-ray and SZ-selected catalogs support no significant difference in morphological bias of these detection techniques. The amount of structure that we see in clusters at different redshifts is consistent with no redshift evolution.

Our study extends Weissmann by including more SPT-detected clusters, and by splitting high-redshift clusters into multiple bins - they had them all in one bin. However, we did not analyze a local sample.

Chapter 5

Conclusions

This dissertation presents an analysis of clusters X-ray surface brightness for the most massive distant clusters drawn from both X-ray and Sunyaev-Zeldovich catalogs.

In Chapter 2 we analyzed the gas density radial distributions and showed that they exhibit a high degree of similarity beyond their central regions in agreement with self-similar models of structure formation. We measured the variations of gas density normalizations around the universal value. Also, we introduced a unifying framework for many functional forms of radial profiles used in the literature.

In Chapter 3 we introduced a new technique that quantifies a cluster's deviation from a relaxed state. We showed that this new substructure measure, $A_{\rm phot}$, is more sensitive to the presence of substructure than its competitors, and is highly insensitive to observational signal-to-noise ratio.

In Chapter 4 we used this new technique as well as other substructure measures to search for evidence of morphological bias between X-ray and SZ-selected clusters and evolution of clusters substructure with redshift. Our results show that the presence of cluster's

substructure is largely independent from cluster's redshift or detection method.

Future work on density profiles should include the comparison of profiles in the cluster's outskirts as observed by *Chandra* and *Suzaku* with the distributions obtained in simulations. After cross-calibration of various observations and simulations, the universal density profile should be extended to larger radii.

The future work on the amount of substructure should include a more detailed (and extended to lower redshift) analysis of the evolution of substructure both in observed and simulated clusters to explain the paradoxical finding that the amount of substructure in clusters doesn't depend on redshift.

Bibliography

Abell, G. 1958, The Astrophysical Journal Supplement Series, 211

Allen, S. W., Evrard, A. E., & Mantz, A. B. 2011, ARA&A, 49, 409

Ameglio, S., Borgani, S., Pierpaoli, E., & Dolag, K. 2007, Monthly Notices of the Royal Astronomical Society, 382, 397

Ameglio, S., Borgani, S., Pierpaoli, E., Dolag, K., Ettori, S., & Morandi, A. 2009, Monthly Notices of the Royal Astronomical Society, 394, 479

Andersson, K., Benson, B. A., Ade, P. A. R., & Aird, K. A. 2011, Arxiv preprint arXiv:

Andersson, K., Peterson, J. R., Madejski, G., & Goobar, a. 2009, The Astrophysical Journal, 696, 1029

Andrade-Santos, F., Lima Neto, G. B., & Laganá, T. F. 2012, ApJ, 746, 139

Angulo, R. E., Springel, V., White, S. D. M., Jenkins, a., Baugh, C. M., & Frenk, C. S. 2012, Monthly Notices of the Royal Astronomical Society, 426, 2046

Arnaud, M., Aghanim, N., & Neumann, D. 2002, Astronomy & Astrophysics, 389, 1

Arnaud, M., Pratt, G. W., Piffaretti, R., Boehringer, H., Croston, J. H., & Pointecouteau, E. 2010, Astronomy and Astrophysics, 517, A92

Benson, B., et al. 2014, in preparation

Bertchinger. 1985, The Astrophysical Journal Supplement Series, 58, 39

Biviano, A. 2000, in Constructing the Universe with Clusters of Galaxies, arXiv:astro-ph/0010409

Blanton, E. L., Randall, S. W., Clarke, T. E., Sarazin, C. L., McNamara, B. R., Douglass, E. M., & McDonald, M. 2011, ApJ, 737, 99

Böhringer, H., et al. 2010, A&A, 514, A32

Böhringer, H., Soucail, G., Mellier, Y., Ikebe, Y., & Schuecker, P. 2000, A&A, 353, 124

Bohringer, H., et al. 2000, The Astrophysical Journal Supplement Series, 129, 435

Böhringer, H., & Werner, N. 2010, A&A Rev., 18, 127

Bonamente, M., et al. 2012, New Journal of Physics, 14, 025010

Bonamente, M., Joy, M. K., Carlstrom, J. E., Reese, E. D., & Laroque, S. J. 2004, ApJ, 614, 56

Bulbul, G. E., Hasler, N., Bonamente, M., & Joy, M. 2010, ApJ, 720, 1038

Buote, D. A. 2002, in Astrophysics and Space Science Library, Vol. 272, Merging Processes in Galaxy Clusters, ed. L. Feretti, I. M. Gioia, & G. Giovannini, 79–107

Buote, D. A., & Humphrey, P. J. 2012a, MNRAS, 420, 1693

——. 2012b, MNRAS, 421, 1399

Buote, D. A., & Tsai, J. C. 1995, ApJ, 452, 522

——. 1996, ApJ, 458, 27

Burenin, R. A., Vikhlinin, A., Hornstrup, A., Ebeling, H., Quintana, H., & Mescheryakov, A. 2007, ApJS, 172, 561

Cassano, R., Ettori, S., Giacintucci, S., Brunetti, G., Markevitch, M., Venturi, T., & Gitti, M. 2010, ApJ, 721, L82

Conselice, C. J. 2003, ApJS, 147, 1

Croston, J. H., Pratt, G. W., Böhringer, H., Arnaud, M., Pointecouteau, E., Ponman, T. J., & Sanderson, A. J. R. 2008, Astronomy & Astrophysics, 487, 431

Cruddace, R., Voges, W., Bo, H., Yentis, D., Schuecker, P., Ebeling, H., & Grandi, S. D. 2002, 239

Doob, J. L. 1949, The Annals of Mathematical Statistics, 20, 393

Duffy, A. R., Schaye, J., Kay, S. T., & Dalla Vecchia, C. 2008, Monthly Notices of the Royal Astronomical Society: Letters, 390, L64

Ebeling, H., White, D. A., & Rangarajan, F. V. N. 2006, MNRAS, 368, 65

Eckert, D., Molendi, S., Gastaldello, F., & Rossetti, M. 2011, Astronomy & Astrophysics, 133, 2

Einasto, J. 1965, Trudy Astrofizicheskogo Instituta Alma-Ata, 5, 87

Evrard, A. E., Mohr, J. J., Fabricant, D., & Geller, M. J. 1993, ApJ, 419, L9

Fabian, A. C., Crawford, C. S., Edge, A. C., & Mushotzky, R. F. 1994, MNRAS, 267, 779

Ferrari, C., Govoni, F., Schindler, S., Bykov, A. M., & Rephaeli, Y. 2008, Space Sci. Rev., 134, 93

Fillmore, J., & Goldreich, P. 1984, ApJ, 281, 1

Fruscione, A., et al. 2006, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 6270, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series

Gao, L., Navarro, J. F., Cole, S., Frenk, C. S., White, S. D. M., Springel, V., Jenkins, A., & Neto, A. F. 2008, Monthly Notices of the Royal Astronomical Society, 387, 536

Giodini, S., Lovisari, L., Pointecouteau, E., Ettori, S., Reiprich, T. H., & Hoekstra, H. 2013, Space Science Reviews, 177, 247

Gonzalez, A. H., Zaritsky, D., & Zabludoff, A. I. 2007, The Astrophysical Journal, 666, 147

Hallman, E. J., & Jeltema, T. E. 2011, MNRAS, 418, 2467

Hallman, E. J., Skillman, S. W., Jeltema, T. E., Smith, B. D., O'Shea, B. W., Burns, J. O., & Norman, M. L. 2010, ApJ, 725, 1053

Harrison, E. 1970, Physical Review D, 1, 2726

Hart, B. C. 2008, PhD thesis, Proquest Dissertations And Theses 2008. Section 0030, Part 0538 163 pages; [Ph.D. dissertation].United States – California: University of California, Irvine; 2008. Publication Number: AAT 3296261. Source: DAI-B 69/01, Jul 2008

High, F. W., et al. 2012, The Astrophysical Journal, 758, 68

Hoekstra, H. 2007, Monthly Notices of the Royal Astronomical Society, 379, 317

Hogg, D. W., Baldry, I. K., Blanton, M. R., & Eisenstein, D. J. 2002, ArXiv Astrophysics e-prints

Hudson, D. S., Mittal, R., Reiprich, T. H., Nulsen, P. E. J., Andernach, H., & Sarazin, C. L. 2010, A&A, 513, A37

Israel, H., et al. 2010, Astronomy & Astrophysics, 58

Jeltema, T. E., Canizares, C. R., Bautz, M. W., & Buote, D. A. 2005, ApJ, 624, 606

Jing, Y. P., Mo, H., Borner, G., & Fang, L. 1995, Mon. Not. R. Astron. Soc., 276, 417

Jing, Y. P., & Suto, Y. 2002, The Astrophysical Journal, 574, 538

Kaiser, N. 1986, MNRAS, 222, 323

Kaiser, N. 1991, ApJ, 383, 104

Kasun, S. F., & Evrard, A. E. 2005, ApJ, 629, 781

Kawahara, H. 2010, The Astrophysical Journal, 719, 1926

Kay, S. T., Powell, L. C., Liddle, A. R., & Thomas, P. A. 2008, MNRAS, 386, 2110

Koester, B. P., et al. 2007, ApJ, 660, 239

Konrad, S., Majer, C. L., Meyer, S., Sarli, E., & Bartelmann, M. 2013, A&A, 553, A118

Krause, E., Pierpaoli, E., Dolag, K., & Borgani, S. 2012, MNRAS, 419, 1766

Kravtsov, A. V., & Borgani, S. 2012, Annual Review of Astronomy and Astrophysics, 50, 353

Kravtsov, A. V., Nagai, D., & Vikhlinin, A. a. 2005, The Astrophysical Journal, 625, 588

Kravtsov, A. V., Vikhlinin, A., & Nagai, D. 2006, asdf

Lacey, C., & Cole, S. 1993, Mon. Not. R. Astron. Soc., 262, 627

Lau, E. T., Nagai, D., Kravtsov, A. V., Vikhlinin, A., & Zentner, A. R. 2012, The Astrophysical Journal, 755, 116

Lau, E. T., Nagai, D., Kravtsov, A. V., & Zentner, A. R. 2011, The Astrophysical Journal, 734, 93

Lynden-Bell, D. 1067, Mon. Not. R. Astron. Soc., 136, 101

Mahdavi, A., & Chang, W. 2011, The Astrophysical Journal, 735, L4

Mahdavi, A., Hoekstra, H., Babul, A., Bildfell, C., Jeltema, T., & Henry, J. P. 2013, The Astrophysical Journal, 767, 116

Mahdavi, a., Hoekstra, H., Babul, a., & Henry, J. P. 2008, Monthly Notices of the Royal Astronomical Society, 384, 1567

Mahdavi, A., Hoekstra, H., Babul, A., Sievers, J., Myers, S. T., & Henry, J. P. 2007, The Astrophysical Journal, 162

Mann, A. W., & Ebeling, H. 2012, Monthly Notices of the Royal Astronomical Society, 420, 2120

Markevitch, M., & Vikhlinin, A. 2007, Phys. Rep., 443, 1

Maughan, B. J., Giles, P. a., Randall, S. W., Jones, C., & Forman, W. R. 2012, Monthly Notices of the Royal Astronomical Society, 421, 1583

McDonald, M. 2011, The Astrophysical Journal, 742, L35

McDonald, M., et al. 2013, ApJ, 774, 23

McNamara, B. R., & Nulsen, P. E. J. 2007, ARA&A, 45, 117

Menanteau, F., et al. 2012, The Astrophysical Journal, 748, 7

Merritt, D., Graham, A. W., Moore, B., Diemand, J., & Terzić, B. 2006, AJ, 132, 2685

Mittal, R., Hicks, A., Reiprich, T. H., & Jaritz, V. 2011, Astronomy & Astrophysics, 133, 1

Mohr, J. J., Evrard, A. E., Fabricant, D. G., & Geller, M. J. 1995, ApJ, 447, 8

Mohr, J. J., Fabricant, D. G., & Geller, M. J. 1993, ApJ, 413, 492

Mohr, J. J., Mathiesen, B., & Evrard, A. E. 1999, ApJ, 517, 627

Moore, B., Quinn, T., Governato, F., Stadel, J., & Lake, G. 1999, Monthly Notices of the Royal Astronomical Society, 310, 1147

Morandi, A., et al. 2012, Mon. Not. R. astr. Soc.

Nagai, D., Kravtsov, A. V., & Vikhlinin, A. 2007, The Astrophysical Journal, 668, 1

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1995, The Astrophysical Journal, 462, 22

Navarro, J. F., et al. 2004, Monthly Notices of the Royal Astronomical Society, 349, 1039

Nelson, K., Rudd, D. H., Shaw, L., & Nagai, D. 2012, The Astrophysical Journal, 751, 121

Neumann, D. M., & Arnaud, M. 1999, Astronomy & Astrophysics, 348, 711

Neumann, D. M., & Bohringer, H. 1997, MNRAS, 289, 123

Noh, Y., & Cohn, J. D. 2012, MNRAS, 426, 1829

Nurgaliev, D., McDonald, M., Benson, B. A., Miller, E. D., Stubbs, C. W., & Vikhlinin, A. 2013, ApJ, 779, 112

O'Hara, T. B., Mohr, J. J., Bialek, J. J., & Evrard, A. E. 2006, ApJ, 639, 64

Peterson, J. R., & Fabian, A. C. 2006, Phys. Rep., 427, 1

Pinkney, J., Roettiger, K., Burns, J. O., & Bird, C. M. 1996, ApJS, 104, 1

Planck Collaboration, et al. 2011, A&A, 536, A14

——. 2013a, ArXiv e-prints, 1303.5089

——. 2013b, A&A, 550, A132

——. 2013c, A&A, 550, A130

——. 2012, A&A, 543, A102

Pointecouteau, E., Arnaud, M., Kaastra, J., & Plaa, J. D. 2004, Astronomy & Astrophysics, 423, 33

Poole, G. B., Babul, A., McCarthy, I. G., Fardal, M. a., Bildfell, C. J., Quinn, T., & Mahdavi, A. 2007, Monthly Notices of the Royal Astronomical Society, 380, 437

Poole, G. B., Fardal, M. A., Babul, A., McCarthy, I. G., Quinn, T., & Wadsley, J. 2006, MNRAS, 373, 881

Pratt, G. W., & Arnaud, M. 2002, A&A, 394, 375

Press, W., & Schechter, P. 1974, The Astrophysical Journal, 187, 425

Rasia, E. 2013, The Astronomical Review, 8, 010000

Reblinsky, K., & Bartelmann, M. 1999, ArXiv Astrophysics e-prints

Richstone, D., Loeb, A., & Turner, E. 1992, ApJ, 393, 447

Rozo, E., et al. 2009, The Astrophysical Journal, 699, 768

Saliwanchik, B. R., et al. 2013, ArXiv e-prints, 1312.3015

Samuele, R., McNamara, B. R., Vikhlinin, a., & Mullis, C. R. 2011, The Astrophysical Journal, 731, 31

Sanderson, A. J. R., Edge, A. C., & Smith, G. P. 2009, MNRAS, 398, 1698

Santos, J. S., Rosati, P., Tozzi, P., Böhringer, H., Ettori, S., & Bignamini, A. 2008, A&A, 483, 35

Santos, J. S., Tozzi, P., Rosati, P., & Böhringer, H. 2010, A&A, 521, A64

Schuecker, P., Boehringer, H., Reiprich, T. H., & Feretti, L. 2001, Astronomy & Astrophysics, 427, 408

Semler, D. R., et al. 2012, ApJ, 761, 183

Sereno, M. 2007, MNRAS, 380, 1207

Shaw, L. D., Nagai, D., Bhattacharya, S., & Lau, E. T. 2010, The Astrophysical Journal, 725, 1452

Sheth, R. 2008, Mon. Not. R. Astron. Soc., 345, 1200

Silva, J. B., Lima, M., & Jr, L. S. 2013, Mon. Not. R. astr. Soc., 436, 2616

Sunyaev, R., & Zeldovich, Y. 1972, Comments on Astrophysics and Space Physics, 4, 173

Vanderlinde, K., et al. 2010, ApJ, 722, 1180

Vazza, F., Brunetti, G., Gheller, C., Brunino, R., & Brüggen, M. 2011, A&A, 529, A17

Ventimiglia, D. A., Voit, G. M., Donahue, M., & Ameglio, S. 2008, ApJ, 685, 118

Vikhlinin, A., et al. 2009, The Astrophysical Journal, 692, 1033

Vikhlinin, A., Burenin, R., Forman, W. R., Jones, C., Hornstrup, A., Murray, S. S., & Quintana, H. 2007, in Heating versus Cooling in Galaxies and Clusters of Galaxies, ed. H. Böhringer, G. W. Pratt, A. Finoguenov, & P. Schuecker, 48, astro-ph/0611438

Vikhlinin, A., et al. 2009a, ApJ, 692, 1033

Vikhlinin, A., Kravtsov, A., Forman, W., Jones, C., Markevitch, M., Murray, S. S., & Speybroeck, L. V. 2006, The Astrophysical Journal, 640, 691

Vikhlinin, A., et al. 2009b, ApJ, 692, 1060

Voit, G. M. 2005, Reviews of Modern Physics, 77, 207

——. 2011, ApJ, 740, 28

Watson, G. 1961, Biometrika, 48, 109

Weißmann, A., Böhringer, H., & Chon, G. 2013a, A&A, 555, A147

Weißmann, A., Böhringer, H., Šuhada, R., & Ameglio, S. 2013b, A&A, 549, A19

Zeldovich, Y. B. 1972, MNRAS, 160, 1P

Zhang, Y.-Y., et al. 2010, The Astrophysical Journal, 711, 1033